

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
28 June 2001 (28.06.2001)

PCT

(10) International Publication Number
WO 01/46408 A2

(51) International Patent Classification⁷: C12N 15/11, 15/86, 15/63, 5/10, 5/22, 1/21, A61K 39/21

(21) International Application Number: PCT/US00/34985

(22) International Filing Date: 22 December 2000 (22.12.2000)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data: 60/173,036 23 December 1999 (23.12.1999) US

(71) Applicant (for all designated States except US): THE GOVERNMENT OF THE UNITED STATES OF AMERICA, as represented by THE SECRETARY, DEPARTMENT OF HEALTH AND HUMAN SERVICES [US/US]; Office of Technology Transfer, National Institutes of Health, 6011 Executive Blvd., Suite 325, Rockville, MD 20852 (US).

(72) Inventor; and

(75) Inventor/Applicant (for US only): PAVLAKIS, George, N. [US/US]; 9 Purdue Court, Rockville, MD 20850 (US).

(74) Agents: FEILER, William, S. et al.; Morgan & Finnegan, L.L.P., 345 Park Avenue, New York, NY 10154 (US).

(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

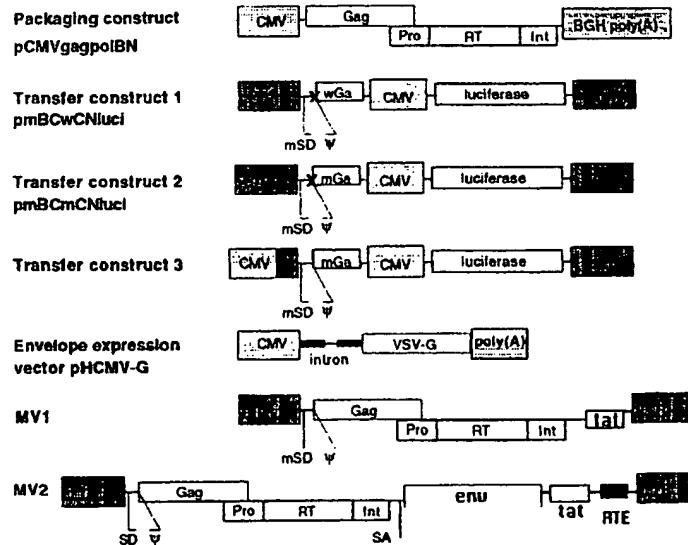
— Without international search report and to be republished upon receipt of that report.

[Continued on next page]

(54) Title: MOLECULAR CLONES WITH MUTATED HIV GAG/POL, SIV GAG AND SIV ENV GENES



WO 01/46408 A2



(57) Abstract: Nucleic acid constructs containing HIV-1 *GAG/POL* and SIV *gag* or SIV *env* genes which have been mutated to remove or reduce inhibitory/instability sequences are disclosed. Viral particles and host cells containing these constructs and/or viral particles are also disclosed. The exemplified constructs and viral particles of the invention may be useful in gene therapy for numerous disorders, including HIV infection, or as a vaccine for HIV-1 immunotherapy and immunoprophylaxis.

WO 01/46408 A2



For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

**MOLECULAR CLONES WITH MUTATED
HIV GAG/POL, SIV GAG AND SIV ENV GENES**

I. TECHNICAL FIELD

The invention relates to nucleic acids comprising mutated HIV-1 *gag/pol* and SIV *gag* gene sequences which are capable of being expressed independently of any SIV or HIV regulatory factors. The invention also relates to nucleic acids comprising a mutated SIV *env* gene sequence, which is capable of being expressed independently of any SIV or HIV regulatory factors. The preferred nucleic acids of the invention are capable of producing infectious viral particles.

The invention also relates to vectors, vector systems and host cells comprising the mutated HIV-1 *gag*, HIV-1 *pol* and/or SIV *gag* gene sequences. The invention also relates host cells comprising these nucleic acids and/or vectors or vector systems. The invention also relates to the use of these nucleic acids, vectors, vector systems and/or host cells for use in gene therapy or as vaccines.

II. BACKGROUND

Until recently, gene therapy protocols have often relied on vectors derived from retroviruses, such as murine leukemia virus (MLV). These vectors are useful because the genes they transduce are integrated into the genome of the target cells, a desirable feature for long-term expression. However, these retroviral vectors can only transduce dividing cells, which limits their use for *in vivo* gene transfer in nonproliferating cells, such as hepatocytes, myofibers, hematopoietic stem cells, and neurons.

Lentiviruses are a type of retrovirus that can infect both dividing and nondividing cells. They have proven extremely efficient at providing long-term gene expression (for up to 6 months) in a variety of nondividing cells (such as, neurons and macrophages) in animal models. See, e.g., Amado et al., *Science* 285:674-676 (July 1999). It has been proposed that the optimal gene transfer system would include a vector based on HIV, or other lentivirus, that can integrate

- 2 -

into the genome of nonproliferating cells. Because retroviruses integrate in the genome of the target cells, repeated transduction is unnecessary. Therefore, in contrast to an adenoviral vector capable of *in vivo* gene delivery, problems linked to the humoral response to injected viral antigens can be avoided. See, e.g., Naldini et al., *Science*, 272:263-267 (1996), p. 263.

HIV and other lentiviruses have a complex genome that, in addition to the essential structural genes (*env*, *gag*, and *pol*), contains regulatory (*tat* and *rev*) and accessory genes (*vpr*, *vif*, *vpu*, and *nef*). HIV has evolved to efficiently infect and express its genes in human cells, and is able to infect nondividing cells such as macrophages because its preintegration complex can traverse the intact membrane of the nucleus in the target cell. This complex contains, in addition to the viral DNA, the enzyme integrase, the product of the *vpr* gene, and a protein encoded by the *gag* gene called matrix. The matrix protein enables the preintegration complex to pass into the nucleus to access the host DNA. Lentiviruses cannot efficiently transduce truly quiescent cells (cells in the G₀ state). However, unlike murine retroviral vectors, in addition to being able to infect dividing cells, HIV-based vectors can achieve effective and sustained transduction and expression of therapeutic genes in nondividing cells, such as hematopoietic stem cells and in terminally differentiated cells such as neurons, retinal photoreceptors, muscle, and liver cells. See, e.g., Amado et al. (July 1999) and Klimatcheva et al., *Frontiers in Bioscience* 4:d481-496 (June 1999), and the references cited therein.

Although lentiviral vectors can be efficient gene delivery vehicles, there are safety concerns due to their origin. Therefore, the field has turned its attention to the development of vectors and production systems with built-in safety features to prevent the emergence of replication competent lentivirus (RCL). For example, in most laboratory applications, lentiviral vectors are generally created in a transient system in which a cell line is transfected with three separate constructs: a packaging construct, a transfer construct, and an envelope encoding construct. The packaging construct contains the elements necessary for vector packaging (except for *env*) and the enzymes required to generate vector particles. The transfer construct contains genetic cis-acting sequences necessary for the vector to infect the target cell and for transfer of the therapeutic (or reporter) gene. The lentivirus *env*

- 3 -

gene is generally deleted from the packaging construct and instead the envelope gene of a different virus is supplied in a third vector "the env-coding vector", although the lentiviruses *env* gene may be used if it is desired that the vector be intended to infect CD4⁺T cells. A commonly used envelope gene is that encoding the G glycoprotein of the vesicular stomatitis virus (VSV-G), which can infect a wide variety of cells and in addition confers stability to the particle and permits the vector to be concentrated to high titers (see, e.g., Naldini et al., Science 272:263-267 (1996) and Akkina et al. J. Virol. 70:2581 (1996). The use of three separate constructs and the absence of overlapping sequences between them minimizes the possibility of recombination during lentivirus (transfer) vector production. In addition, because no viral proteins are expressed by the lentiviral (transfer) vector itself, they do not trigger an effective immune response against cells expressing vector in animal models (a particular problem with vectors based on adenovirus). See, e.g., Amado et al., Science 285:674-676 (July 1999) and the references cited therein. See also Naldini et al. Science 272:263-267 (1996).

The initial packaging plasmids contained most HIV genes except for *env*. In an effort to improve safety, subsequent HIV vectors have been produced in which the packaging plasmid is devoid of all accessory genes. This process does not interfere with efficient vector production and significantly increases the safety of the system because potential RCLs lack the accessory genes necessary for efficient replication of HIV in humans. Although these vectors can transduce growth-arrested cell lines and neurons *in vivo*, they have been reported to not efficiently transduce macrophages. The accessory gene *vpr* is believed to be necessary for HIV infection of these cells using these HIV vectors. See, Zufferey et al., Nature Biotechnol. 15:871-875 (1997). In contrast, as discussed later herein, the HIV-based lentiviral vectors of the present invention do not need any HIV accessory genes in order to be able to infect human macrophages and the other cells tested.

The requirement of *vpr* or *vif* for efficient transduction of liver cells has also been reported. See, e.g., Kafri et al., Nature Genet. 17:314 (1997). These results indicate that the requirement of accessory genes for efficient lentivirus-mediated gene transfer is dependent on the type of cell chosen as target, suggesting

- 4 -

that future applications of lentiviral vectors may involve vector constructs with different accessory genes, as needed.

Zufferey et al., (1997) describe an HIV vector system in which the virulence genes, *env*, *vif*, *vpr*, *vpu*, and *nef* have been deleted. This multiply attenuated vector conserved the ability to transduce growth-arrested cells and monocyte-derived macrophages in culture, and could efficiently deliver genes in vivo into adult neurons. The packaging plasmids described Zufferey et al. (1997) and Naldini et al. (1996) encode Rev and Tat, in addition to Gag and Pol.

Lentiviral vectors engineered to become packaged into virions in the absence of the regulatory gene *tat* have also been described. See, e.g., Kim et al., J. Virol. 72:811-816 (1998) and Miyoshi et al. J. Virol. 72:8150-8157 (1998). In these vectors the *tat* gene has been removed from the packaging plasmid. Kim et al. state that *tat* is not necessary as long as the serial 5' LTR promoter is replaced with a strong constitutive promoter. It also has other advantages for HIV therapy. Replacement of the HIV-1 LTR with a constitutive HCMV promoter permits the use of anti-Tat molecules such as Tat transdominant mutants or Tat activation response element decoys as therapeutic agents, since they will not affect vector production. (see p. 814, col. 2). The removal of the *tat* gene eliminates an essential virulence factor that could contribute to a possible RCL. Kim et al. (1998) describe a vector system which does not contain *tat*, *vif*, *vpr*, *vpu* and *nef*. The preferred vector system includes the *rev* gene which, the authors state "with RRE, is required for efficient RNA handling in this system." (p. 811, col. 2). However, Kim et al. also constructed Rev independent constructs using CTE. Kim et al. state that the *rev*/RRE components could be removed by using a sequence such as the Mason-Pfizer monkey virus (MPMV) constitutive transport element (CTE), thereby eliminating all accessory proteins, but this leads to a significant reduction in titer.

Srinivasakumar et al., J. Virol. 71:5841-5848 (1997) describes the generation of stable HIV-1 packaging lines that constitutively express high levels of HIV-1 structural proteins in either a Rev-dependent or a Rev-independent fashion. These cell lines were used to assess gene transfer by using a HIV-1 vector expressing the hygromycin B resistance gene and to study the effects of Rev, Tat, and Nef on the vector titer. The Rev-independent cell lines were created by using

- 5 -

gag-pol and *env* expression vectors that contain the MPMV CTE. This article describes the construction of four plasmids, among others: CMV gagpol-RRE and pCMVenv, which require Rev coexpression for HIV-1 structural gene expression, and pCMV gagpol-CTE and pCMVenv-CTE, which do not. To create Rev-containing and Rev-independent packaging, cell lines, CMT3 cells were transfected with vectors expressing Gag, Gag-Pol, and Env, using a calcium phosphate transfection procedure.

By creating an HIV vector which contained the MPMV CTE (pTR167-CTE) and a packaging cell line which expressed the HIV structural proteins in a Rev-independent fashion, the authors were able to obtain a HIV vector system that functions completely without Rev. The titer of the vector obtained from this system was essentially the same as that obtained from a parallel system which contained Rev. The authors state that, in this context, the CTE seemed to substitute completely for Rev-RRE functions, similar to what was previously observed in transient-expression assays with Rev-dependent constructs. This is in contrast to situations where several rounds of HIV replication were measured. In those cases, titers from CTE-containing viruses were always reduced by at least 1 log unit compared to viruses utilizing Rev and the RRE. (See, Srinivasakumar et al., p. 5847).

The authors state that the advantages of having a HIV vector system that works in the absence of Rev opens the possibility of using it as a delivery vehicle for intracellular immunization against Rev function. Genes encoding Rev antagonists that have dramatic inhibitory effects on HIV replication, such as Rev M10 or RRE decoys, could be introduced into an HIV vector and put into cells normally infectable by HIV. Expression of the "anti-Rev" gene would be expected to dampen HIV infection. Any residual HIV replication should lead to activation of the vector LTR (by Tat) and create a vector-derived RNA that would be packaged by proteins derived from the infectious virus. In this scenario, the wild-type virus would act as a helper that may allow the spread of vector particles to previously nonimmunized cells. Because of the additional vector spread, it is likely that this type of scheme will be more effective in modulating HIV infection *in vivo* than one

- 6 -

based on traditional retrovirus vectors. The authors state that they are currently testing this approach in model systems. (See, Srinivasakumar et al., p. 5847).

Another development in the quest for a safe system is the so-called self-inactivating (SIN) vector. See, e.g., Yu et al., Proc Natl Acad Sci USA 83:3194-8 (1986) and Miyoshi et al., J. Virol. 72:8150 (1998). In Yu et al., a retrovirus-derived vector SIN vector was designed for the transduction of whole genes into mammalian cells. The SIN vector of Yu et al. contains a deletion of 299 base pairs in the 3' long terminal repeat (LTR), which includes sequences encoding the enhancer and promoter functions. When viruses derived from such vectors were used to infect NIH 3T3 cells, the deletion was transferred to the 5' LTR, resulting in the transcriptional inactivation of the provirus in the infected cell. Introduction of a hybrid gene (human metallothionein-promoted c-fos) into cells via a SIN vector was not associated with rearrangements and led to the formation of an authentic mRNA transcript, which in some cases was induced by cadmium. The vector described in Miyoshi et al. also contains a deletion the 3' (downstream) LTR. A sequence within the upstream LTR serves as a promoter under which the viral genome is expressed. The deletion introduced in the downstream LTR is transferred to the upstream LTR during reverse transcription. This deletion inactivates the LTR promoter and eliminates the production of vector RNA. The gene (or genes) to be transferred (e.g., a reporter or therapeutic gene) is expressed from an exogenous viral or cellular promoter that is inserted into the lentivirus vector. An important safety feature of SIN vectors is that inactivation of the promoter activity of the LTR reduces the possibility of insertional mutagenesis (of the transfer vector) into the host genome. In addition, because the expression of the (transfer) vector RNA is eliminated, the potential for RCL production in the target cell is further minimized. SIN vectors should be particularly useful in gene transfer experiments designed to study the regulated expression of genes in mammalian cells. Absence of enhancer and promoter sequences in both LTRs of the integrated provirus should also minimize the possibility of activating cellular oncogenes and may provide a safer alternative to be used in human gene therapy. Other modifications to enhance safety and specificity include the use of specific internal promoters that regulate gene expression, either temporally or with tissue or cell specificity.

- 7 -

Other strategies to improve safety in human studies would be to use nonhuman lentiviruses such as simian immunodeficiency virus, bovine immunodeficiency virus, or equine infectious anemia virus. Of these, vectors derived from the feline immunodeficiency virus have been engineered to efficiently transduce nondividing human cells. See, e.g., Poeschla et al., Nature Med. 4:354-357 (1998) and WO 99/15641. In addition, White et al., J. Virol. 73:2832-2840 (April 1999) described lentiviral vectors using human and simian immunodeficient virus elements in attempt to improve safety by reducing the likelihood of recombination between packaging constructs and transfer constructs.

The development of efficient packaging lines has proven challenging because expression of the VSV-G envelope and a number of HIV proteins is toxic to cells. Recently, a producer line has been designed in which the expression of packaging genes and VSV-G, and therefore the production of vector, can be turned on at will. Kafri et al., J. Virol. 73:576-584 (1999). The cell line can be expanded for scale-up vector production when the expression of toxic genes is turned off. This cell line produces high titer vector without generating RCL. Hematopoietic stem cells transduced with an HIV vector were transplanted into rhesus macaques as described by Donahue et al. Blood 92 (suppl. 1), abstract 4648.5 (1998) with at least a 14-month follow-up. At that time the procedure proved to be safe; all animals in the study have remained healthy without evidence of circulating HIV or vector. See, Amado et al., Science 285:674-676 (July 1999).

Many gene therapy protocols have been designed to correct a number of inherited metabolic, infectious, or malignant diseases using the hematopoietic stem cell. This cell has the capacity to self-renew and to differentiate into all of the mature cells of the blood and immune systems. Many diseases that affect these systems could potentially be treated by the stable introduction of therapeutic genes into stem cells. Recently, lentiviral vectors were shown to bypass the need for *ex vivo* stem cell stimulation (which is necessary when using murine retroviral vectors), by mediating efficient gene transfer into very primitive human stem cells that contributed to stable, long-term reconstitution of SCID mouse bone marrow with many hematopoietic lineages. See, e.g., Miyoshi et al., Science 283:682 (1999). Similarly, in a rhesus macaque model of autologous transplantation with

- 8 -

lentivirus-transduced stem cells, multilineage gene expression was found, suggesting transduction of an early blood cell progenitor under conditions of minimal stem cell stimulation, ordinarily insufficient for transduction with murine retroviruses. See, Donahue et al., Blood 92 (suppl. 1), abstract 4648.5 (1999) and Amado et al., Science 285:674-676 (July 1999).

In HIV infection, another advantage of lentiviral vectors designed against HIV is their potential to be mobilized by HIV in the infected patient, because the virus supplies all of the necessary elements for packaging of the vector. If these mobilized vectors contained the HIV envelope, they could efficiently transfer their genes (for example, genes custom-designed to confer resistance against HIV) into CD4⁺ T cells, protecting them from subsequent HIV infection. Lentiviral vectors can also be designed to efficiently express their genes only in CD4⁺ T cells that are infected with HIV (so called *tat*-inducible vectors). In these vectors, all HIV genes, including *tat* and *rev*, are ablated; cis-acting sequences required for integration, expression, and packaging are retained, and expression is dependent on the activity of the HIV LTR (which requires transactivation by Tat). It has been shown that in this system, vector expression is induced efficiently upon HIV infection. Moreover, in the absence of genes that confer resistance against HIV, stable integration of this vector in permissive cell lines resulted in inhibition of HIV replication. Although the mechanism of HIV inhibition has not been completely elucidated, preliminary results suggest that this vector competes with HIV at the level of reverse transcription. See, An et al., J. Virol., in press, and Amado et al., Science 285:674-676 (1999).

A number of other potential medical applications, where the modification of the genetic material of quiescent cells could result in the prevention or reversal of a disease process, are beginning to be explored. For example, the finding that lentiviral vectors can mediate stable and long-term gene transfer by direct injection of vector into the rat and mouse retina has lent support to the notion of gene therapy for the treatment of retinitis pigmentosa. This degenerative disease of the retina is characterized by photoreceptor cell death, resulting in a slow progression to blindness. Mutations in the cGMP phosphodiesterase β subunit

- 9 -

(PDE β) gene of rod photoreceptors lead to an autosomal recessive form of retinitis pigmentosa in humans, and in the *rd* mouse model of the disease. Previous studies have shown that adenovirus and adeno-associated virus-mediated PDEP subretinal gene transfer results in a delay in photoreceptor cell death. Using the *rd* mouse model, a recent study demonstrated that photoreceptors could be rescued in up to 50% of eyes injected with a lentivirus vector containing the murine PDE β gene. In contrast with the short-term expression previously obtained with adenovirus vectors, PDE β expression in this study persisted for at least 24 weeks. This finding points to the potential success of gene therapy in a disease that currently lacks effective treatment. See, Takahashi et al., J. Virol., 73:7812-7816 (Sept. 1999) and Amado et al. Science, 285:674-676 (1999).

In nature, the expression of *gag*, *pol*, and *env* of HIV-1 depends on the presence of the viral Rev protein. This dependence is, at least in part, due to the presence of negatively acting sequences (inhibitory or instability elements [INS]) located within unspliced and partially spliced mRNAs. The positive interaction of Rev with the Rev-responsive element [RRE] in these mRNAs counteracts the negative effects of the inhibitory sequences.

None of the above references teach or suggest that the *gag* and/or *pol* genes described therein may be replaced with the *gag* and/or *pol* genes in which the inhibitory/instability have been mutated to render their expression Rev-independent. Furthermore, there is no disclosure of the specific HIV-1 *gag/pol* or SIV *gag* mutated genes described herein.

The *gag/pol* clone of the invention was made using the method for eliminating inhibitory/instability regions from a gene as first described in U.S. patent application Serial No. 07/858,747, filed March 27, 1992, (inventors, G. Pavlakis and B. Felber) entitled "Method of Eliminating Inhibitory/Instability Regions from mRNA" and later described in a Continuation-in-Part ("CIP") application, filed as PCT application PCT/US93/02908 on March 29, 1993 and U.S. Patent Nos. 5,972,596 and 5,965,726. The disclosure of the CIP application was published as International Publication No. WO 93/20212 on October 14, 1993. (The disclosures of these patents and patent applications are specifically

- 10 -

incorporated by reference herein in their entirety.) The method was also described in Schwartz et al., J. Virol. 66:7176-7182 (1992).

Schneider et al., J. Virol. 71:4892-4903 (1997), extend the work described in the patent applications and in Schwartz et al. by identifying and characterizing additional INS within *gag*, *protease* and *pol* genes and mutating them in a similar manner. Schneider et al. disclose nucleic acid constructs which contain completely mutated HIV-1 *gag* genes, but only partially mutated HIV-1 *pol* genes.

Schneider et al. demonstrate that expression vectors containing an intact or nearly intact p55^{gag} region allow the production of immature viral particles in mammalian cells in the absence of any other HIV proteins. The introduction of additional mutations in the *protease* region allowed efficient production of Gag/protease, which resulted in processing of the Pr55^{gag} precursor and production of mature Gag particles with a lentivirus-like conical-core structure.

Schneider et al. disclose that Rev-independent expression vectors allow the efficient expression of Gag proteins in many cell lines that are not able to support efficient Rev-RRE-dependent rescue of these RNAs. Schneider et al. also disclose that *gag/pol* expression vectors may be important for vaccination approaches against HIV-1, since the *gag/pol* region is more conserved than is the *env* region and may be important for an effective immune response against HIV and for protection against infection. They also state that efficient HIV gene expression in many cells is also of interest for possible gene transfer experiments using lentiviral vectors in nondividing or slowly dividing cells, since HIV and the other lentiviruses are able to infect quiescent cells.

Pavlakis et al., Natl Conf Hum Retroviruses Relat Infect (2nd). (1995), 91, state that Rev-independent Gag expression vectors were able to produce viral particles in human and mouse cells in the absence of any other HIV proteins, and that additional mutations in the *pol* region allowed the expression of the protease and the processing of the p55 gag precursor. Direct DNA injection of TAT and Rev independent Gag expression vectors in mouse muscle resulted in Gag expression detected by ELISA and in anti-gag antibody response. Several Rev-and Tat- independent Gag expression cassettes were inserted into retroviral vectors and

- 11 -

cell lines expressing Gag or Gag fragments that are dominant negative inhibitors of HIV-1 were constructed.

Shiver et al. (1996) describe the results of DNA vaccination of mice and non-human primates with mutated plasmid DNA encoding either mutated genes encoding HIV-1 gag (p55 gag) or env (gp120 or gp160). Both *gag* and *env* vaccine recipients exhibited antigen-specific cytotoxic and helper T lymphocyte (CTL, Th) responses. The results are stated to demonstrate that DNA vaccines elicited long-lived T cell responses in both mice and nonhuman primates that were disseminated throughout the lymphatics.

III. SUMMARY OF THE INVENTION

The invention relates to nucleic acids comprising the nucleic acid sequence of the mutated HIV-1 *gag/pol* gene shown in Figure 1 (SEQUENCE ID NO:1) and vectors and vector systems comprising these nucleic acids.

The invention also relates to nucleic acids comprising the nucleic acid sequence of the mutated SIV *gag* gene shown in Figure 3 and vectors and vector systems comprising these nucleic acids.

The invention also relates to nucleic acids comprising the mutated SIV *env* gene shown in Figure 17 and vectors and vector systems comprising these nucleic acids.

The invention also relates to products produced by the nucleic acids, e.g., mRNA, protein, and infectious viral particles.

The invention also relates to compositions comprising these nucleic acids and/or their expression products.

The invention also relates to host cells comprising these nucleic acids, vector systems or viral particles.

The invention also relates to uses of these nucleic acids, vector systems, host cells, expression products, and/or compositions to produce mRNA, proteins, and/or infectious viral particles, and/or to induce antibodies and/or cytotoxic or helper T lymphocytes.

The invention also relates to the use of these nucleic acid constructs, vectors, vector systems and or host cells for use in immunotherapy and

- 12 -

immunoprophylaxis, e.g., as a vaccine, or in genetic therapy after expression, preferably in humans. The nucleic acid constructs of the invention can include or be incorporated into lentiviral vectors or other expression vectors or they may also be directly injected into tissue cells resulting in efficient expression of the encoded protein or protein fragment. These constructs may also be used for *in-vivo* or *in-vitro* gene replacement, e.g., by homologous recombination with a target gene *in-situ*.

IV. BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1. DNA sequence of a mutated HIV-1 *gag/pol* molecular clone (SEQUENCE ID NO:1). The gagpol terminator is located at positions 4305-4397 of SEQUENCE ID NO:1.

Fig. 2. Comparison of the sequence of the wild -type and mutated *pol* region in pCMVgagpolBNkan. Position #1 in the figure is position 2641 in plasmid pCMVgagpolBNkan. The comparison starts at position 1872 from the gag initiator ATG.

Fig. 3. DNA sequence of a mutated SIV *gag* molecular clone (SIVgagDX).

Fig. 4. Comparison of the mutated SIV *gag* DNA sequence in SIVgagDX with the wild type SIV sequence from Simian (macaque) immunodeficiency virus isolate 239, clone lambda siv 239-1 (GenBank accession No. M33262).

Fig. 5. Schematic diagram of some components of sample versions of a lentiviral system. BGH poly (A): bovine growth hormone poly (A) signal; MSD: mutated splice donor site; ψ: encapsidation signal; SD, splice donor site; SA, splice acceptor site; "X" indicates that the ATG codon of the partial *gag* gene sequence is mutated so that translation of this gene does not occur.

Fig. 6. Schematic diagram of the packaging construct pCMVgagpolBNkan.

Fig. 7. Schematic diagram of transfer construct 1: pmBCwCNluci. The packaging signal, the CMV promoter and the coding region for the luciferase gene are flanked by the 5' and 3' HIV-1 LTRs, which provide promoter and

- 13 -

Polyadenylation signals, as indicated by the arrows. Three consecutive arrows indicate the U5, R, and U3 regions of the LTR, respectively. The transcribed portions of the LTRs are shown in black. Some restriction endonuclease cleavage sites are also indicated.

Fig. 8. Schematic diagram of transfer construct 1: pmBCmCNluci. Symbols are as above.

Fig. 9. DNA sequence of packaging construct pCMVgagpolBNkan.

Fig. 10. DNA sequence of transfer construct 1: pmBCwCNluci.

Fig. 11. DNA sequence of transfer construct 1: pmBCmCNluci.

Figure 12:

Fig. 12. Nucleotide sequence of the region BssHII (711) to ClaI (830) in wild-type HIV-1 molecular clones HXB2 and NL4-3, and in the transfer constructs. The translation initiator signal for Gag protein (ATG) is underlined. pmBCwCNluci and pmBCmCNluci (transfer constructs 1 and 2) contain the sequence mBCwCN. Transfer construct 3 contains the sequence m2BCwCN. In contrast to the sequence mBCwCN, m2BCwCN has different mutations at the 5' splice site region and has an intact Gag ATG.

Fig. 13. Bar graph showing levels of gag protein that is released from cells upon transient transfection with pCMVgagpolBNkan (labeled pCMVBNKan in the figure).

Fig. 14. Bar graph showing reverse transcriptase activity from the Rev-independent gag-pol HIV-1 vector pCMVgagpolBNkan (labeled pCMVBNKan in the figure).

Fig. 15. Bar graphs showing the amount of luciferase per nanogram of p24 Gag protein detected in cells transduced with PCMVgagpolBNkan Rev-independent gag-HIV-1 based retroviral vectors. The results show that with PCMVgagpolBNkan Rev-independent gag-HIV-1 based retroviral vectors display high transduction efficiency in (A) 293 cells, (B) human lymphoid cells, (C) human myeloid cells (U937), as well as (D) non-dividing cells such as primary human macrophages.

Fig. 16. Schematic diagram of the SIV envelope encoding vector CMVkan/R-R-SIVenvCTE.

- 14 -

Fig. 17. DNA sequence of the SIV envelope encoding vector CMVkan/R-R-SIVenvCTE containing a mutated SIV env gene.

V. MODES FOR CARRYING OUT THE INVENTION

It is to be understood that both the foregoing general description and the following detailed description are exemplary and explanatory only, and are not restrictive of the invention, as claimed. The accompanying drawings, which are incorporated in and constitute a part of the specification, illustrate an embodiment of the invention and, together with the description, serve to explain the principles of the invention.

One aspect of the invention comprises vectors that encode the Gag and/or Pol of HIV-1 in a Rev-independent manner. An example of such a vector which is described herein is the plasmid pCMVgagpolBNkan, which encodes the complete Gag and Pol of HIV-1 in a Rev-independent manner, and also contains a gene conferring kanamycin resistance. This plasmid is Tat and Rev-independent and was generated by eliminating the inhibitory/instability sequences present in the *gag/pol* mRNA without altering the amino acid sequence of the proteins coded by the genes.

The *gag/pol* clone of the invention is a DNA construct of the *gag/pol* region of HIV which has had the inhibitory/instability regions removed. The construct is expected to be useful as a component a new type of lentivirus vector for use in gene therapy or as a vaccine.

The *gag*, *pol* or *gag/pol* sequences of the invention can be highly expressed in human and other mammalian cells in the absence of any other regulatory and structural protein of HIV, including Rev. When the *gag/pol* sequences are combined with a sequence encoding an envelope protein, such as the VSV G protein or the HIV envelope protein (e.g., in the same vector or in another expression vector), infectious virus is produced after transfection into human cells. When a gene encoding a non-HIV envelope protein is used, for example, in the presence of the HIV *gag/pol* gene, the virus particles produced would contain only the HIV proteins Gag and Pol.

- 15 -

Lentiviral vectors or vector systems based on the *gag*, *pol* or *gag/pol* sequences of this invention, as exemplified by the Rev-independent pCMVgagpol BNkan construct described herein, may be used for gene therapy *in vivo* (e.g., parenteral inoculation of high titer vector) or *ex vivo* (e.g., *in vitro* transduction of patient's cells followed by reinfusion into the patient of the transduced cells). These procedures are been already used in different approved gene therapy protocols.

The HIV *gag/pol* clone and SIV *gag* clone of the invention were made using the method for eliminating inhibitory/instability regions from a gene as described in U.S. patent application Serial No. 07/858,747, filed March 27, 1992, and also in related U.S. Patent Nos. 5,972,596 and 5,965,726, which are incorporated by reference herein. This method does not require the identification of the exact location or knowledge of the mechanism of function of the INS. Generally, the mutations are such that the amino acid sequence encoded by the mRNA is unchanged, although conservative and non-conservative amino acid substitutions are also envisioned where the protein encoded by the mutated gene is substantially similar to the protein encoded by the non-mutated gene. The mutated genes can be synthetic (e.g., synthesized by chemical synthesis), semi-synthetic (e.g., a combination of genomic DNA, cDNA, or PCR amplified DNA and synthetic DNA), or recombinantly produced. The genes also may optionally not contain introns. The nucleic acids of the invention may also contain Rev-independent fragments of these genes which retain the desired function (e.g., for antigenicity of Gag or Pol, particle formation (Gag) or enzymatic activity (Pol)), or they may also contain Rev-independent variants which have been mutated so that the encoded protein loses a function that is unwanted in certain circumstances. In the latter case, for example, the gene may be modified to encode mutations (at the amino acid level) in the active site of reverse transcriptase or integrase proteins to prevent reverse transcription or integration. Rev-independent fragments of the *gag* gene are described in U.S. patent application Serial No. 07/858,747, filed March 27, 1992, and also in related U.S. Patent Nos. 5,972,596 and 5,965,726, which are incorporated by reference herein.

In addition to being capable of producing HIV Gag and Pol proteins in the absence of Rev regulatory protein in a cell *in vivo*, the HIV *gag/pol* clone and

- 16 -

SIV *gag* clone of the invention are also capable of producing HIV Gag and Pol proteins in the absence of any added cis acting transport element, such as CTE or CTE-like elements (collectively referred herein as RNA Transport Elements (RTE)). Experiments indicate that the mutated vectors of the invention for SIV gag are far superior to those adding CTE (see Qiu et al., J Virol. 73:9145-52 (1999)).

The expression of the proteins encoded by these vectors after transfection into human cells may be monitored at both the level of RNA and protein production. RNA levels are quantitated by methods known in the art, e.g., Northern blots, S1 mapping or PCR methods. Protein levels may also be quantitated by methods known in the art, e.g., western blot or ELISA or fluorescent detection methods. A fast non-radioactive ELISA protocol can be used to detect gag protein (DUPONT or COULTER gag antigen capture assay).

At least three types of lentiviral vectors based on the *gag/pol* genes of the invention for use in gene therapy and/or as a vaccine are envisioned, i.e., lentiviral vectors having

- a) no round of replication (i.e., a zero replication system)
- b) one round of replication
- c) a fully replicating system

For a system with no round of replication, a *gag/pol* gene, or separate *gag* and *pol* genes, or fragments of these genes, expressed using appropriate transcription units, e.g., a CMV promoter and a BGH poly (A) site. This will allow expression of the *gag/pol* unit (or *gag* or *pol* or fragment(s) thereof) for vaccine purposes. This expression can be accomplished without the production of any functional retroviral enzymes, provided that the appropriate mutation(s), e.g., a missense mutation, are introduced. In a zero replication system, a virus stock will be administered to the cells or animals of interest. For example, if one creates and uses a virus stock with the exemplified system using the packaging vector PCMVgagpolBNKan, the transfer construct pmBCwCNluci or pmBCmCNluci, and the envelope containing vector pHCMV-G, one obtains a zero replication system. The virus particles produced by such system can infect cells, and the reverse transcribed transfer construct DNA will go into the nucleus but, because the coding regions for viral structural proteins are not present, there will be no virus expression.

- 17 -

and replication (0 rounds). If one transfets cells *in vivo* with the same 3 DNAs, they will go to the nucleus, express viral proteins, make infectious virus particles and go out and infect another cell or cells (1 round). Since *in vivo* delivery of three plasmids may result in lower expression, at least two different embodiments are envisioned. In the first, two plasmids may be used, e.g., MV1 shown in Fig. 5 and an envelope expression plasmid such as pHCMV-G. Other plasmids encoding functional envelopes from HIV, SIV, or other retroviruses can also be used. Transfection by the two plasmids results in infectious virus that can infect and integrate into new cells (1 round). The infected cells produce gagpol but virus propagation is not possible in the absence of env.

For a system with one round of replication, at least two additional embodiments are envisioned. In the first method, a combination of the genes, e.g., a *gag/pol* gene, an *env* encoding gene and, preferably, a gene encoding a reporter protein or other polynucleotide or protein of interest, are delivered into the cells of interest *in vivo*. As discussed above for the exemplified system, if one transfets cells *in vivo* with the same 3 DNAs, they will go to the nucleus, express viral proteins, make infectious virus particles, be released and infect another cell or cells (1 round).

In another embodiment, the same result (i.e., only one round of replication) can be obtained by using transfer vectors that have deletions in the 3' LTR and in which a heterologous-promoter (e.g., the CMV-promoter, or inducible promoter, or tissue-specific promoter), is used in place of the '3'LTR promoter. The mutations in the 3' LTR making it inactive upon reverse transcription and integration. This is because the integrated provirus derives both its 5' LTR and its 3' LTR from the 3' LTR of the starting (transfer) construct. (This is a well-known property of all retroviruses and has been used to make self-inactivating vectors (SIN)). There are several reasons one may want to inactivate the incoming LTR promoter, one of which is to use a different tissue specific or regulated promoter for expression of a gene of interest in the integrated provirus. Note that, with SIN vectors, if one uses a viral stock made *in vitro* after transfection into cells and collection of infectious virus, there will be no round of replication. If one transfets cells with the DNAs *in vivo*, there will be one round of replication. If functional

- 18 -

gag, *pol*, or *env* are not included in the DNA mix, there will not be any infection at all (i.e., infectious viruses will not be made).

A fully replicating Rev-independent system has not been constructed yet, although it is expected that a functional system can be constructed using Rev-independent *gag/pol* and *env* sequences. If desired, extra posttranscriptional control elements such as the CTE element, which can replace Rev and give infectious virus (see e.g., Zolotukhin et al., J.Virol.68:944-7952 (1994)) are included. The fully replicating system should be in one piece, containing the LTR, packaging signal, *gag/pol*, splice site, *env*, *tat*, one or more CTE or CTE-like elements (if desired for optimal results), and LTR. *Tat* is thought to be required in this construct, at least in non-permissive cells. Such a system is depicted in Figure 5, (construct MV2). In this system, a cell or animal of interest (preferably human) would be infected with virus stock that then propagates. CTE or CTE-like elements (depicted in construct MV2 as RTE (RNA Transport Elements)) are desirable since they have been shown to improve expression, and since many retroviruses require the presence of posttranscriptional control elements. There are several types of CTE and CTE-like elements, and these elements appear to work via a different pathway from the Rev-RRE pathway. See, e.g., Tabernero et al., J Virol. 71:95-101 (1997). See also, Pavlakis and Nappi, PCT/US99/11082, filed May 22, 1999, published as WO 99/61596 on December 2, 1999 (and incorporated herein by reference), which describes a new type of post-transcriptional control element that is able to replace CTE and HIV RRE/Rev. The Pavlakis-Nappi element does not work in the same way as CTE and does not have any sequence or structure homology.

In a preferred embodiment, a lentiviral system of the invention comprises the following three components:

1. a packaging vector containing nucleic acid sequences encoding the elements necessary for vector packaging such as structural proteins (except for HIV *env*) and the enzymes required to generate vector particles, the packaging vector comprising at least a mutated HIV or SIV *gag/pol* gene of the invention;

- 19 -

2. a transfer vector containing genetic *cis*-acting sequences necessary for the vector to infect the target cell and for transfer of the therapeutic or reporter or other gene(s) of interest, the transfer vector comprising the encapsidation signal and the gene(s) of interest or a cloning site for inserting the gene(s) of interest;
- and
3. a vector containing sequences encoding an element necessary for targeting the viral particle to the intended recipient cell, preferably the gene encoding the G glycoprotein of the vesicular stomatitis virus (VSV-G) or amphotrophic MuLV or lentiviral *envs*.

Using the CMV promoter or other strong, high efficiency, promoter instead of the HIV-1 LTR promoter in the packaging vector, high expression of *gag*, *pol*, or *gag/pol* can be achieved in the total absence of any other viral protein. The exchange of the HIV-1 LTR promoter with other promoters is beneficial in the packaging vector or other vectors if constitutive expression is desirable and also for expression in other mammalian cells, such as mouse cells, in which the HIV-1 promoter is weak. Vectors containing the sequences of the invention can be used for the Rev independent production of HIV-1 Gag/Pol, HIV-1 Gag, HIV-1 Pol, and SIV Gag proteins. In certain embodiments, the presence of heterologous promoters will also be desired in the transfer vector and the envelope encoding vector, when such vectors are used.

The gene(s) of interest are chosen according to the effect sought to be achieved. For gene therapy purposes there will be at least one therapeutic gene encoding a gene product which is active against the condition it is desired to treat or prevent. Alternatively or additionally, there may be a gene which acts as a marker by encoding a detectable product. Therapeutic genes may encode, for example, an anti-sense RNA, a ribozyme, a transdominant negative mutant of a target protein, a toxin, a conditional toxin, an antigen that induces antibodies or helper T-cells or cytotoxic T-cells, a single chain antibody or a tumor suppresser protein. See, e.g., WO 98/17816.

- 20 -

An even more extensive list of genes of interest for use in lentiviral vectors is described, e.g., in WO 99/04026 on page 10, line 20 to page 12, line 7. Table 2 of Klimatcheva et al. (1999) also provides a list of disorders and target cells for gene therapy, as well as a number of lentiviral vectors used by others. This list includes genetic/metabolic deficiencies, viral infection and cancer. Inherited genetic defects such as adenosine deaminase deficiency, familial hypercholesterolemia, cystic fibrosis, mucopolysaccharidosis type VII, types I and II diabetes, classical phenylketonuria and Gaucher disease are diseases which are listed as being possible to overcome by lentiviral vector-mediated gene therapy because they constitute single-gene deficiencies for which the involved genes are known. Viral diseases are also listed as constituting appropriate targets for lentiviral gene delivery. In particular, a number of gene therapy approaches have been proposed for the treatment of HIV infection and, for some of these strategies, phase I studies have recently begun in humans. The article states that preliminary studies have dealt with defective murine oncoviruses for delivery of anti-sense RNAs, ribozymes and trans-dominant proteins against HIV replication.

In any of the vectors, but preferably in the transfer vector, an inserted gene could have an internal ribosomal entry site (IRES), e.g., from picornaviral RNA. An IRES will be used in circumstances that one wants to express two proteins from the same promoter. For example one protein of interest and a marker gene, e.g., green fluorescent protein (GFP) or a marker gene and a drug resistance gene (e.g. the firefly luciferase gene and neomycin phosphotransferase gene) as described on p. 58 of WO 99/04026, for example. Using an IRES the expression of the two proteins is coordinated. A further gene or genes may also be present under the control of a separate promoter. Such a gene may encode for example a selectable marker, or a further therapeutic agent which may be among the therapeutic agents listed above. Expression of this gene may be constitutive; in the case of a selectable marker this may be useful for selecting successfully transfected packaging cells, or packaging cells which are producing particularly high titers of the retroviral vector particles. Alternatively or additionally, the selectable marker may be useful for selecting cells which have been successfully infected with the lentiviral vector and have the provirus integrated into their own genome.

- 21 -

One way of performing gene therapy is to extract cells from a patient, infect the extracted cells with a lentiviral vector and reintroduce the cells back into the patient. A selectable marker may be used to provide a means for enriching for infected or transduced cells or positively selecting for only those cells which have been infected or transduced, before reintroducing the cells into the patient. This procedure may increase the chances of success of the therapy. Selectable markers may be for instance drug resistance genes, metabolic enzyme genes, or any other selectable markers known in the art. Typical selection genes encode proteins that confer resistance to antibiotics and other toxic substances, e.g., histidinol, puromycin, hygromycin, neomycin, methotrexate etc. and cell surface markers.

However, it will be evident that for many gene therapy applications of lentiviral vectors, selection for expression of a marker gene may not be possible or necessary. Indeed expression of a selection marker, while convenient for *in vitro* studies, could be deleterious *in vivo* because of the inappropriate induction of cytotoxic T lymphocytes (CTLs) directed against the foreign marker protein. Also, it is possible that for *in vivo* applications, vectors without any internal promoters will be preferable. The presence of internal promoters can affect for example the transduction titres obtainable from a packaging cell line and the stability of the integrated vector. Thus, single transcription unit vectors, which may be bi-cistronic or poly-cistronic, coding for one or two or more therapeutic genes, may be the preferred vector designed for use *in vivo*. See, e.g., WO 98/17816.

Suitable host or producer cells for use in the invention are well known in the art. Many lentiviruses have already been split into replication defective genomes and packaging components. For those which have not the technology is available for doing so. The producer cell encodes the viral components not encoded by the vector genome such as the Gag, Pol and Env proteins. The *gag*, *pol* and *env* genes may be introduced into the producer cell transiently, or may be stably integrated into the cell genome to give a packaging cell line. The lentiviral vector genome is then introduced into the packaging cell line by transfection or transduction to create a stable cell line that has all of the DNA sequences required to produce a lentiviral vector particle. Another approach is to introduce the different

- 22 -

DNA sequences that are required to produce lentiviral vector particle, e.g., the *env* coding construct, the *gag-pol* coding construct and the transfer construct into the cell simultaneously by transient triple transfection.

Target cells identified by Klimatcheva et al. (1999), and the references cited therein, include airway epithelial cells for cystic fibrosis; retinal photoreceptor cells for retinitis pigmentosa; progenitors for red blood cells, macrophages, and lymphocytes for hematopoietic disorders, sickle cell anemia, β-thalassemia, lysosomal storage disorders, mucopolysaccharidoses, and severe combined immunodeficiency syndrome; bone marrow cells and macrophages for Gaucher's disease; liver cells for familial hypercholesterolaemia; T-lymphocytes and macrophages for HIV infection; brain tissue, neurons, and glial cells for neurodegenerative diseases such as Parkinson's and Alzheimer's diseases; endothelial cells and cardiac myocytes for cardiovascular diseases; and cancer cells in various tissues (e.g. liver or brain) for cancer. Target cells for other diseases would be apparent to one of skill in the art.

Vaccines and pharmaceutical compositions comprising at least one of the nucleic acid sequences, vectors, vector systems, or transduced or transfected host cells of the invention and a physiologically acceptable carrier are also part of the invention.

As used herein, the term "transduction" generally refers to the transfer of genetic material into the host via infection, e.g., in this case by the lentiviral vector. The term "transfection" generally refers to the transfer of isolated genetic material into cells via the use of specific transfection agents (e.g., calcium phosphate, DEAE Dextran, lipid formulations, gold particles, and other microparticles) that cross the cytoplasmic membrane and deliver some of the genetic material into the cell nucleus.

Systems similar to those described herein can be produced using elements of lentiviruses in addition to the HIV and/or SIV genes described herein.

Pharmaceutical Compositions

The pharmaceutical compositions of the invention contain a pharmaceutically and/or therapeutically effective amount of at least one nucleic acid

- 23 -

construct, vector, vector system, viral particle/virus stock, or host cell (i.e., agents) of the invention. In one embodiment of the invention, the effective amount of an agent of the invention per unit dose is an amount sufficient to cause the detectable expression of the gene of interest. In another embodiment of the invention, the effective amount of agent per unit dose is an amount sufficient to prevent, treat or protect against deleterious effects (including severity, duration, or extent of symptoms) of the condition being treated. The effective amount of agent per unit dose depends, among other things, on the species of mammal inoculated, the body weight of the mammal and the chosen inoculation regimen, as is well known in the art. The dosage of the therapeutic agents which will be most suitable for prophylaxis or treatment will also vary with the form of administration, the particular agent chosen and the physiological characteristics of the particular patient under treatment. The dose is administered at least once. Subsequent doses may be administered as indicated.

To monitor the response of individuals administered the compositions of the invention, mRNA or protein expression levels may be determined. In many instances it will be sufficient to assess the expression level in serum or plasma obtained from such an individual. Decisions as to whether to administer another dose or to change the amount of the composition administered to the individual may be at least partially based on the expression levels.

The term "unit dose" as it pertains to the inocula refers to physically discrete units suitable as unitary dosages for mammals, each unit containing a predetermined quantity of active material (e.g., nucleic acid, virus stock or host cell) calculated to produce the desired effect in association with the required diluent. The titers of the virus stocks to be administered to a cell or animal will depend on the application and on type of delivery (e.g., *in vivo* or *ex vivo*). The virus stocks can be concentrated using methods such as centrifugation. The titers to be administered *ex vivo* are preferably in the range of 0.001 to 1 infectious unit /cell. Another method of generating viral stocks is to cocultivate stable cell lines expressing the virus with the target cells. This method has been used to achieve better results when using traditional retroviral vectors because the cells can be infected over a longer period of time and they have the chance to be infected with multiple copies of the vector.

- 24 -

For *in vivo* administration of nucleic acid constructs, vectors, vector systems, virus stocks, or cells which have been transduced or transfected *ex vivo*, the dose is to be determined by dose escalation, with the upper dose being limited by the onset of unacceptable adverse effects. Preliminary starting doses may be extrapolated from experiments using lentiviral vectors in animal models, by methods known in the art, or may be extrapolated from comparisons with known retroviral (e.g., adenoviral) doses. Generally, small dosages will be used initially and, if necessary, will be increased by small increments until the optimum effect under the circumstances is reached. Exemplary dosages are within the range of 10^8 up to approximately 5×10^{15} particles.

Inocula are typically prepared as a solution in a physiologically acceptable carrier such as saline, phosphate-buffered saline and the like to form an aqueous pharmaceutical composition.

The agents of the invention are generally administered with a physiologically acceptable carrier or vehicle therefor. A physiologically acceptable carrier is one that does not cause an adverse physical reaction upon administration and one in which the nucleic acids are sufficiently soluble to retain their activity to deliver a pharmaceutically or therapeutically effective amount of the compound. The pharmaceutically or therapeutically effective amount and method of administration of an agent of the invention may vary based on the individual patient, the indication being treated and other criteria evident to one of ordinary skill in the art. A therapeutically effective amount of a nucleic acid of the invention is one sufficient to prevent, or attenuate the severity, extent or duration of the deleterious effects of the condition being treated without causing significant adverse side effects. The route(s) of administration useful in a particular application are apparent to one or ordinary skill in the art.

Routes of administration of the agents of the invention include, but are not limited to, parenteral, and direct injection into an affected site. Parenteral routes of administration include but are not limited to intravenous, intramuscular, intraperitoneal and subcutaneous. The route of administration of the agents of the invention is typically parenteral and is preferably into the bone marrow, into the CSF intramuscular, subcutaneous, intradermal, intraocular, intracranial, intranasal,

- 25 -

and the like. See, e.g., WO 99/04026 for examples of formulations and routes of administration.

The present invention includes compositions of the agents described above, suitable for parenteral administration including, but not limited to, pharmaceutically acceptable sterile isotonic solutions. Such solutions include, but are not limited to, saline and phosphate buffered saline for nasal, intravenous, intramuscular, intraperitoneal, subcutaneous or direct injection into a joint or other area.

In providing the agents of the present invention to a recipient mammal, preferably a human, the dosage administered will vary depending upon such factors as the mammal's age, weight, height, sex, general medical condition, previous medical history and the like.

The administration of the pharmaceutical compositions of the invention may be for either "prophylactic" or "therapeutic" purpose. When provided prophylactically, the compositions are provided in advance of any symptom. The prophylactic administration of the composition serves to prevent or ameliorate any subsequent deleterious effects (including severity, duration, or extent of symptoms) of the condition being treated. When provided therapeutically, the composition is provided at (or shortly after) the onset of a symptom of the condition being treated.

For all therapeutic, prophylactic and diagnostic uses, one or more of the agents of the invention, as well as antibodies and other necessary reagents and appropriate devices and accessories, may be provided in kit form so as to be readily available and easily used.

Where immunoassays are involved, such kits may contain a solid support, such as a membrane (e.g., nitrocellulose), a bead, sphere, test tube, rod, and so forth, to which a receptor such as an antibody specific for the target molecule will bind. Such kits can also include a second receptor, such as a labeled antibody. Such kits can be used for sandwich assays to detect toxins. Kits for competitive assays are also envisioned.

- 26 -

VI. INDUSTRIAL APPLICABILITY

Mutated genes of this invention can be expressed in the native host cell or organism or in a different cell or organism. The mutated genes can be introduced into a vector such as a plasmid, cosmid, phage, virus or minichromosome and inserted into a host cell or organism by methods well known in the art. In general, the mutated genes or constructs containing these mutated genes can be utilized in any cell, either eukaryotic or prokaryotic, including mammalian cells (e.g., human (e.g., HeLa), monkey (e.g., Cos), rabbit (e.g., rabbit reticulocytes), rat, hamster (e.g., CHO and baby hamster kidney cells) or mouse cells (e.g., L cells), plant cells, yeast cells, insect cells or bacterial cells (e.g., E. coli). The vectors which can be utilized to clone and/or express these mutated genes are the vectors which are capable of replicating and/or expressing the mutated genes in the host cell in which the mutated genes are desired to be replicated and/or expressed. See, e.g., F. Ausubel et al., Current Protocols in Molecular Biology, Greene Publishing Associates and Wiley-Interscience (1992) and Sambrook et al. (1989) for examples of appropriate vectors for various types of host cells. The native promoters for such genes can be replaced with strong promoters compatible with the host into which the gene is inserted. These promoters may be inducible. The host cells containing these mutated genes can be used to express large amounts of the protein useful in enzyme preparations, pharmaceuticals, diagnostic reagents, vaccines and therapeutics.

Mutated genes or constructs containing the mutated genes may also be used for in-vivo or in-vitro gene therapy. For example, a mutated gene of the invention will produce an mRNA in situ to ultimately increase the amount of protein expressed. Such gene include viral genes and/or cellular genes. Such a mutated gene is expected to be useful, for example, in the development of a vaccine and/or genetic therapy.

The constructs and/or proteins made by using constructs encoding the mutated gag, env, and pol genes could be used, for example, in the production of diagnostic reagents, vaccines and therapies for AIDS and AIDS related diseases. The inhibitory/instability elements in the HIV-1 gag gene may be involved in the

- 27 -

establishment of a state of low virus production in the host. HIV-1 and the other lentiviruses cause chronic active infections that are not cleared by the immune system. It is possible that complete removal of the inhibitory/instability sequence elements from the lentiviral genome would result in constitutive expression. This could prevent the virus from establishing a latent infection and escaping immune system surveillance. The success in increasing expression of the entire *gag/pol* gene by eliminating the inhibitory sequence element suggests that one could produce lentiviruses without any negative elements. Such lentiviruses could provide a novel approach towards attenuated vaccines.

For example, vectors expressing high levels of Gag can be used in immunotherapy and immunoprophylaxis, after expression in humans. Such vectors include retroviral vectors and also include direct injection of DNA into muscle cells or other receptive cells, resulting in the efficient expression of gag, using the technology described, for example, in Wolff et al., Science 247:1465-1468 (1990), Wolff et al., Human Molecular Genetics 1(6):363-369 (1992) and Ulmer et al., Science 259:1745-1749 (1993). Further, the gag constructs could be used in transdominant inhibition of HIV expression after the introduction into humans. For this application, for example, appropriate vectors or DNA molecules expressing high levels of p55^{gag} or p37^{gag} would be modified to generate transdominant gag mutants, as described, for example, in Trono et al., Cell 59:113-120 (1989). The vectors would be introduced into humans, resulting in the inhibition of HIV production due to the combined mechanisms of gag transdominant inhibition and of immunostimulation by the produced gag protein. In addition, the gag constructs of the invention could be used in the generation of new retroviral vectors based on the expression of lentiviral gag proteins. Lentiviruses have unique characteristics that may allow the targeting and efficient infection of non-dividing cells. Similar applications are expected for vectors expressing high levels of env.

Identification of similar inhibitory/instability elements in SIV indicates that this virus is a convenient model to test these hypotheses. SIV similarly modified could be used in place of HIV in an effort to further minimize the possibility of rearrangement events that would lead to the generation of infectious HIV.

- 28 -

The following examples illustrate certain embodiments of the present invention, but should not be construed as limiting its scope in any way. Certain modifications and variations will be apparent to those skilled in the art from the teachings of the foregoing disclosure and the following examples, and these are intended to be encompassed by the spirit and scope of the invention.

EXAMPLE 1

Rev-Independent HIV-1 Gag/Pol Molecular Clone

Figure 1 shows the DNA sequence of a Rev-independent HIV-1 *gag/pol* molecular clone. This DNA sequence shown encodes the complete Gag and Pol of HIV-1 and can be expressed in a Rev-independent manner when operably linked to a promoter. The Rev-independent *gag* sequence was described in U.S. Patent Nos. 5,972,596 and 5,965,726 and the Rev-independent *pol* sequence was generated by eliminating the inhibitory/instability sequences using the methods described in U.S. Patent Nos. 5,972,596 and 5,965,726. Others have reportedly made Rev independent *gag* sequences by optimizing codon usage for human cells (see, e.g., WO 98/34640).

Figure 2 shows an alignment of the sequence of the wild - type and mutated *pol* region in pCMVgagpolBNkan. Position #1 in the figure is position 2641 in plasmid pCMVgagpolBNkan.

The elimination of INS in *gag*, *pol* and *env* regions allows the expression of high levels of authentic HIV-1 structural proteins in the absence of the Rev regulatory factor of HIV-1.

EXAMPLE 2

Rev-Independent SIV Gag Molecular Clone

Figure 3 shows the DNA sequence of a Rev-independent SIV *gag* molecular clone, SIVgagDX. Figure 4 shows the comparison of wild type (WT) and mutant (SIVgagDX) sequences. The wild type SIV sequence is from Simian (macaque) immunodeficiency virus isolate 239, clone lambda siv 239-1 (GenBank accession No. M33262).

- 29 -

EXAMPLE 3

Rev-Independent SIV Env Molecular Clone

Figure 16 shows a schematic diagram, and figure 17 shows the DNA sequence, of the "env-coding" vector CMVkan/R-R-SIVenvCTE, which is an example of a vector comprising a mutated lentiviral *env* gene sequence which is capable of being expressed independently of any SIV or HIV regulatory factors. "CMV" denotes the cytomegalovirus promoter; "SRV-CTE" denotes the constitutive transport element (CTE) of Simian Retrovirus Type 1; "all-STOP" denotes a sequence providing translational stops in all three reading frames; "BGH terminator" denotes the bovine growth hormone polyadenylation signal. Other posttranscriptional control elements can be used instead of the indicated SRV-CTE, for example the one described by Pavlakis and Nappi, PCT/US99/11082, filed May 22, 1999, which was published as WO 99/61596 on December 2, 1999 (and which is incorporated herein by reference).

As mentioned previously above, such a vector encoding a lentiviral *env* gene may be used if it is desired that the vector infect CD4⁺T cells. Also as mentioned previously above, the CTE element (i.e., the SRV-CTE element in the case of vector CMVkan/R-R-SIVenvCTE), can be replaced with another post-transcriptional control element, such as the Pavlakis-Nappi element, that is able to replace CTE and HIV RRE/Rev. See Pavlakis and Nappi, PCT/US99/11082, filed May 22, 1999, which was published as WO 99/61596 on December 2, 1999 (and which is incorporated herein by reference).

EXAMPLE 4

Lentiviral Vector System

Figure 5 is a schematic of some of the components of a preliminary version of the Rev-independent lentiviral vector system exemplified herein, including a packaging construct and three different transfer vectors which may be used. In the lentiviral system exemplified herein, the packaging construct also contains the gene for kanamycin resistance. The lentiviral system exemplified herein also contains the vector pHCMV-G, which is shown in Figure 5.

- 30 -

In the packaging construct shown in Figure 5, "CMV" denotes the cytomegalovirus promoter, "Gag" denotes the gag gene, which generates components of the virion core, "Pro" denotes "protease" "RT" denotes "reverse transcriptase," "Int" denotes "integrase" and "BGH poly (A)" denotes the bovine growth hormone polyadenylation signal. The protease, reverse transcriptase, and integrase genes comprise the "pol" gene. In transfer construct 1, "LTR" denotes the HIV "long terminal repeat", which contains a HIV promoter; "mSD" denotes "mutated splice donor site," which is present in the construct so that splicing of the RNA transcript does not occur; "ψ" denotes the encapsidation signal; "wGA" denotes part of the wild-type *gag* gene which contains sequences believed to be necessary for encapsidation; "X" indicates that the ATG codon of the partial *gag* gene sequence is mutated so that translation of this gene does not occur; "CMV" denotes the cytomegalovirus promoter and luciferase is used as a reporter gene. Luciferase can be replaced with any gene of interest. Another HIV LTR is present at the 3' end of transfer construct 1. Replacement of this LTR in constructs such as the transfer construct 1, 2, or 3 with a promoter-enhancer deleted HIV LTR leads to inactivation of LTR after integration. Transfer construct 2 is similar to transfer construct 1, the difference being that a mutated part of the *gag* gene (denoted "mGa") is used instead of the wild-type part of the *gag* gene. Transfer construct 3 (pm2BCwCNluci) has different mutations at the 5' splice site and has an intact ATG codon so that translation of part of the mutated *gag* gene occurs. Transfer construct 3 also has a 5' CMV promoter instead of a 5' LTR promoter. This construct is expressed independent of the presence of HIV Tat protein. The transfer constructs expressed from the LTR promoter are partially dependent on Tat protein. In 293 cells significant expression can be achieved in the absence of Tat. See, e.g., Valentin et al., Proc. Natl Acad. Sci. U S A. 95:8886-91 (1988).

EXAMPLE 5

Generation of Packaging Construct pCMVgagpol BNkan

Figure 6 shows a schematic map of the packaging construct pCMVgagpolBNKan. The nucleotide numbering is that of the HXB2R sequence

- 31 -

(Genbank accession number K03455 and M38432), where +1 is the start of transcription.

The sequence in HIV-1 *gag/pol* region was mutated in order to eliminate all the INS. The fragment from the beginning of *gag* to BsrGI site in *pol*, and the fragment KE [KpnI(3700)- EcoRI(4194)] were previously mutated described in Schneider et al., J Virol. 71: 4892-4903 (1997) and in U.S. Patent Nos. 5,972,596 and 5,965,726.

To generate pCMVgagpolBNkan, three fragments within HIV-1 *pol* region were mutated. They are fragment BP [BsrGI(2207)-PflMI(3032)], fragment PK [PflMI(3032)-KpnI(3700)] and fragment EN [EcoRI(4194)-NdeI(4668)]. Mutagenesis was performed using a modified version of the method described by Ho et al., Gene 77: 51-59 (1989) and DNA shuffling (Zhao and Arnold, Nucl. Acid Res. 25(6), 1307-1308 (1997). Sixteen oligonucleotides extending over the complete sequence of the three fragments were designed. Six oligos corresponded to fragment BP, six to fragment PK, and four to fragment EN (the oligonucleotides ranged from 130 to 195 bases in length; adjacent oligos overlapped by twenty nucleotides). Each fragment was assembled in two steps:

1) PCR; the reaction was carried out in standard *pfu* buffer with 10 pmol of each purified big oligo, 0.2 mM of each dNTPs and 2.5 u *pflu* DNA polymerase enzyme (Stratagene) in a 50 μ l final volume. The PCR program was: 3 min 96°C followed by 50 cycles of 1 min 94°C, 1 min 55°C, and 1 min + 5 s/cycle 72°C, ended by 7 min at 72°C. After PCR, the big oligonucleotides were removed from the assembled mutated fragment.

2) The second step was to specifically amplify the assembled products with 30 mer primers located at the 5' and 3' end of each mutated fragment. One microliter of the assembled PCR product was used as template in a 25-cycle PCR reaction with 50 pmol of each primer, 1 x *pflu* buffer, 0.2 mM of each dNTP and 2.5 u *pflu* DNA polymerase in a 50 μ l final volume. The PCR program was: 3 min 96°C, 10 cycles of 30 s 94°C, 30 s 55°C, 45 s 72°C, followed by another 14 cycles of 30 s 94°C, 30 s 55°C, 45 s + 20 s/cycle 72°C, and finally 7 min 72°C. This program gave a single PCR product of the correct size. The amplified BP, PK

- 32 -

and EN fragments were individually cloned into PCR-script™ vector using PCR-script™ Amp SK(+) Cloning Kit (Stratagene). Clones were randomly selected and sequenced. The correct BP, PK and EN fragments together with fragment KE previously mutated by Schneider et al. were ligated between BsrGI and KpnI site of p55AM1-R5 (which was previously described in Schneider et al., J. Virol. 71: 4892-4903 (1997)) to produce a completely mutated *gag/pol* ORF. The new plasmid containing the completely mutated *gag/pol* was named pLTRgagpolBN. BN stands for the modification of the fragment between BsrGI and NdeI. The mutated *gag/pol* was then cloned into a CMVkan vector containing the cytomegalovirus major late promoter (GenBank accession no. X17403) and the kanamycin resistance gene, resulting in pCMVgagpolBNkan. The plasmid backbone comes from pVR1332 provided by Vical Inc., and described in Hartikka et al., Hum Gene Ther. 7:1205-17 (1996).

It is understood that different plasmid backbones can be used, e.g., to provide good expression *in vivo*, in the case of DNA injection, for example.

EXAMPLE 6

Construction of Transfer Vectors pmBCwCNluci and pmBCmCNluci

The HIV-1 sequence BC, between BssHII (257) and Clal (376), contains the major splice donor site and the encapsidation signal. Six oligos (33 to 46 bases) were designed to introduce mutations on the splice donor site and the AUG start codon of *gag*. The BC fragment was assembled, amplified and sequenced as described in the section concerning the construction of pCMVgagpolBN.

The mutated BC fragment and a fragment of wild type *gag* between Clal (376) and Nsi (793) were placed between the BssHII and Nsi sites of p55RRE (Schneider et al., J. Virol. 71:4892-4903 (1997)) to generate pmBCwCN. In parallel, the fragment between Clal (376) and NsiI sites of mutated *gag* from p55BM1-10SD+ was used to generate pmBCmCN. (p55BM1-10SD+ is similar to p55BM1-10, which is described in Schneider et al. (1997), but contains in addition the intact splice donor and encapsidation site upstream of *gag*). The region between

- 33 -

NsiI and XhoI containing 3' part of *gag* and RRE in pmBCwCN and pmBCmCN was replaced by a ClaI-XhoI fragment containing CMV promoter and luciferase gene from pHRI'-CMVluci (vector from D. Trono) to generate pmBCwCNluci and pmBCmCNluci (which are shown as transfer constructs 1 and 2 in Figure 5, and schematically depicted in Figures 7 and 8, respectively). The sequences of these plasmids are shown in Figures 10 and 11, respectively. Different versions of these plasmids have also been created, by standard procedures, with variations in the region of the encapsidation site, the first splice donor site, and the initiator *gag* AUG. For example, the transfer construct pm2BcwCNluci (which is shown as transfer construct 3 in Fig. 5) has different mutations in the 5' splice site region and has an intact ATG. A comparison of the sequences in the BssHII-Cla I region of transfer constructs 1 and 2 (mBCwCN frag), transfer construct 3 (m2BCwCN frag), HXB2 and NL43 is shown in Fig. 12.

EXAMPLE 7

Preparation of Viral Particles

Lentiviral particles were generated by transient cotransfection of 293 human kidney cells with a combination of three plasmids: pCMVgagpolBNkan, pmBCwCNluci or pmBCmCNluci (transfer vector) and pHCMV-G (Yee et al., Proc. Natl. Acad. Sci., USA, 91:9564-9568 (1994) a plasmid coding for the envelope VSV-G (glycoprotein of vesicular stomatitis virus).

The day before the transfection, 293 cells were plated at a density of 10^6 cells/plate on a 60 mm plate. Plasmid DNA was transfected by the Ca-phosphate precipitation method in the following proportions: 3 μ g packaging construct, 6 μ g transfer construct and 100 ng VSV-G encoding construct, pHCMV-G. [Note that the LTR promoter can be expressed in 293 cells in the absence of Tat with a moderate decrease in efficiency. The transfer constructs can be fully Tat independent after replacement of the LTR promoter with a CMV (see, e.g., transfer construct 3 in Fig. 5) or other promoter in such a way that the mRNA start site is at the beginning of the LTR R region.] In the present experiments for preparation of viral particles 500 ng of a Tat expression plasmid was included in the transfection.

- 34 -

Cells were washed the day after transfection and were kept in DMEM medium for another 48 hours before the supernatants were harvested. Supernatants were spun at 1,200 rpm for 7 mins to eliminate any floating cells. pCMVgagpolBNkan produces high levels of Gag protein that is efficiently released from the cells (Figure 13), and also produces high levels of functional Pol as judged by levels of reverse transcriptase activity similar to those found upon expression of complete HIV-1 (Figure 14).

Supernatants from 293 transfected cells were used to transduce several human cell lines (293, Jurkat, U937) and non-dividing human primary macrophages.

EXAMPLE 8

Cell Transduction

Transduction was performed by incubating for 3-4 hours at 37°C the target cells with 1-2 ml of supernatant containing the retroviral vectors. The amount of retroviral vector present in the supernatant was normalized by p24 content (measured by ELISA). Equal amounts of p24 gag protein were used for infection of cells. This way, differences in production of the different preparations was minimized.

The macrophages used for transduction were isolated from the peripheral blood of healthy donors by adherence to plastic. Cells were cultured in RPMI + 20% fetal calf serum (FCS) + 10% human serum (HS). After 1 week, non-adherent cells were washed off with PBS and the macrophages were kept in culture for another 1-2 weeks in the absence of human serum. The cells were washed 2-4 times with PBS before transduction.

Cells were harvested 48 hours after transduction (seven days for primary macrophages) and the transduction efficiency was determined by measuring luciferase activity in cell extracts from the cultures. The results of the transduction experiments in 293 Jurkat, U937 and primary macrophages are shown in Figure 15A-D. These results demonstrate that Rev-independent gag-HIV-1 based retroviral vectors display high transduction efficiency in (A) 293 cells, (B) human

- 35 -

lymphoid cells, (C) human myeloid cells (U937), as well as (D) non-dividing cells such as primary human macrophages.

EXAMPLE 9

Use Of Nucleic Acids of the Invention In Immunoprophylaxis Or Immunotherapy

In postnatal gene therapy, new genetic information has been introduced into tissues by indirect means such as removing target cells from the body, infecting them with viral vectors carrying the new genetic information, and then reimplanting them into the body; or by direct means such as encapsulating formulations of DNA in liposomes; entrapping DNA in proteoliposomes containing viral envelope receptor proteins; calcium phosphate co-precipitating DNA; and coupling DNA to a polylysine-glycoprotein carrier complex. In addition, in vivo infectivity of cloned viral DNA sequences after direct intrahepatic injection with or without formation of calcium phosphate coprecipitates has also been described. mRNA sequences containing elements that enhance stability have also been shown to be efficiently translated in Xenopus laevis embryos, with the use of cationic lipid vesicles. See, e.g., J.A. Wolff, et al., Science 247:1465-1468 (1990) and references cited therein.

Recently, it has also been shown that injection of pure RNA or DNA directly into skeletal muscle results in significant expression of genes within the muscle cells. J.A. Wolff, et al., Science 247:1465-1468 (1990). Forcing RNA or DNA introduced into muscle cells by other means such as by particle-acceleration (N.-S. Yang, et al. Proc. Natl. Acad. Sci. USA 87:9568-9572 (1990); S.R. Williams et al., Proc. Natl. Acad. Sci. USA 88:2726-2730 (1991)) or by viral transduction should also allow the DNA or RNA to be stably maintained and expressed. In the experiments reported in Wolff et al., RNA or DNA vectors were used to express reporter genes in mouse skeletal muscle cells, specifically cells of the quadriceps muscles. Protein expression was readily detected and no special delivery system was required for these effects. Polynucleotide expression was also obtained when the composition and volume of the injection fluid and the method of injection were modified from the described protocol. For example, reporter enzyme activity was

- 36 -

reported to have been observed with 10 to 100 μ l of hypotonic, isotonic, and hypertonic sucrose solutions, Opti-MEM, or sucrose solutions containing 2mM CaCl₂ and also to have been observed when the 10- to 100- μ l injections were performed over 20 min. with a pump instead of within 1 min.

Enzymatic activity from the protein encoded by the reporter gene was also detected in abdominal muscle injected with the RNA or DNA vectors, indicating that other muscles can take up and express polynucleotides. Low amounts of reporter enzyme were also detected in other tissues (liver, spleen, skin, lung, brain and blood) injected with the RNA and DNA vectors. Intramuscularly injected plasmid DNA has also been demonstrated to be stably expressed in non-human primate muscle. S. Jiao et al., Hum. Gene Therapy 3:21-33 (1992).

It has been proposed that the direct transfer of genes into human muscle in situ may have several potential clinical applications. Muscle is potentially a suitable tissue for the heterologous expression of a transgene that would modify disease states in which muscle is not primarily involved, in addition to those in which it is. For example, muscle tissue could be used for the heterologous expression of proteins that can immunize, be secreted in the blood, or clear a circulating toxic metabolite. The use of RNA and a tissue that can be repetitively accessed might be useful for a reversible type of gene transfer, administered much like conventional pharmaceutical treatments. See J.A. Wolff, et al., Science 247:1465-1468 (1990) and S. Jiao et al., Hum. Gene Therapy 3:21-33 (1992).

It had been proposed by J.A. Wolff et al., supra, that the intracellular expression of genes encoding antigens might provide alternative approaches to vaccine development. This hypothesis has been supported by a recent report that plasmid DNA encoding influenza A nucleoprotein injected into the quadriceps of BALB/c mice resulted in the generation of influenza A nucleoprotein-specific cytotoxic T lymphocytes (CTLs) and protection from a subsequent challenge with a heterologous strain of influenza A virus, as measured by decreased viral lung titers, inhibition of mass loss, and increased survival. J. B. Ulmer et al., Science 259:1745-1749 (1993).

- 37 -

Therefore, it appears that the direct injection of RNA or DNA vectors encoding the viral antigen can be used for endogenous expression of the antigen to generate the viral antigen for presentation to the immune system without the need for self-replicating agents or adjuvants, resulting in the generation of antigen-specific CTLs and protection from a subsequent challenge with a homologous or heterologous strain of virus.

CTLs in both mice and humans are capable of recognizing epitopes derived from conserved internal viral proteins and are thought to be important in the immune response against viruses. By recognition of epitopes from conserved viral proteins, CTLs may provide cross-strain protection. CTLs specific for conserved viral antigens can respond to different strains of virus, in contrast to antibodies, which are generally strain-specific.

Thus, direct injection of RNA or DNA encoding the viral antigen has the advantage of being without some of the limitations of direct peptide delivery or viral vectors. See J.A. Ulmer et al., supra, and the discussions and references therein). Furthermore, the generation of high-titer antibodies to expressed proteins after injection of DNA indicates that this may be a facile and effective means of making antibody-based vaccines targeted towards conserved or non-conserved antigens, either separately or in combination with CTL vaccines targeted towards conserved antigens. These may also be used with traditional peptide vaccines, for the generation of combination vaccines. Furthermore, because protein expression is maintained after DNA injection, the persistence of B and T cell memory may be enhanced, thereby engendering long-lived humoral and cell-mediated immunity.

1. Vectors for the immunoprophylaxis or
immunotherapy against HIV-1

The mutated *gag*, *pol* or *gag/pol* sequences will be inserted in expression vectors using a strong constitutive promoter such as CMV or RSV, or an inducible promoter such as HIV-1.

The vector will be introduced into animals or humans in a pharmaceutically acceptable carrier using one of several techniques such as injection of DNA directly into human tissues; electroporation or transfection of the

- 38 -

DNA into primary human cells in culture (*ex vivo*), selection of cells for desired properties and reintroduction of such cells into the body, (said selection can be for the successful homologous recombination of the incoming DNA to an appropriate preselected genomic region); generation of infectious particles containing the *gag* gene, infection of cells *ex vivo* and reintroduction of such cells into the body; or direct infection by said particles *in vivo*.

Substantial levels of protein will be produced leading to an efficient stimulation of the immune system.

In another embodiment of the invention, the described constructs will be modified to express mutated Gag proteins that are unable to participate in virus particle formation. It is expected that such Gag proteins will stimulate the immune system to the same extent as the wild-type Gag protein, but be unable to contribute to increased HIV-1 production. This modification should result in safer vectors for immunotherapy and immunophrophylaxis.

EXAMPLE 10

Inhibition of HIV-1 Expression Using Transdominant (TD)-TD-Gag-TD Rev or Td Gag-Pro-TD Rev Genes

Direct injection of DNA or use of vectors other than retroviral vectors will allow the constitutive high level of trans-dominant Gag (TDgag) in cells. In addition, the approach taken by B.K. Felber et al., *Science* 239:184-187 (1988) will allow the generation of retroviral vectors, e.g. mouse-derived retroviral vectors, encoding HIV-1 TDgag, which will not interfere with the infection of human cells by the retroviral vectors. In the approach of Felber, et al., *supra*, it was shown that fragments of the HIV-1 LTR containing the promoter and part of the polyA signal can be incorporated without detrimental effects within mouse retroviral vectors and remain transcriptionally silent. The presence of Tat protein stimulated transcription from the HIV-1 LTR and resulted in the high level expression of genes linked to the HIV-1 LTR.

The generation of hybrid TDgag-TDRev or TDgag-pro-TDRev genes and the introduction of expression vectors in human cells will allow the efficient production of two proteins that will inhibit HIV-1 expression. The incorporation of

- 39 -

two TD proteins in the same vector is expected to amplify the effects of each one on viral replication. The use of the HIV-1 promoter in a manner similar to one described in B.K. Felber, et al., supra, will allow high level Gag and Rev expression in infected cells. In the absence of infection, expression will be substantially lower. Alternatively, the use of other strong promoters will allow the constitutive expression of such proteins. This approach could be highly beneficial, because of the production of a highly immunogenic gag, which is not able to participate in the production of infectious virus, but which, in fact, antagonizes such production. This can be used as an efficient immunoprophylactic or immunotherapeutic approach against AIDS.

Examples of trans-dominant mutants are described in Trono et al., Cell 59:112-120 (1989).

1. Generation of constructs encoding transdominant Gag mutant proteins

Gag mutant proteins that can act as trans-dominant mutants, as described, for example, in Trono et al., supra, will be generated by modifying vector p37M1-10D or p55M1-13P0 to produce transdominant Gag proteins at high constitutive levels.

The transdominant Gag protein will stimulate the immune system and will inhibit the production of infectious virus, but will not contribute to the production of infectious virus.

The added safety of this approach makes it more acceptable for human application.

- 40 -

VII. REFERENCES

U.S. Patent No. 5,972,596 issued October 26, 1999 (Pavlakis and Felber)

U.S. Patent No. 5,965,726 issued October 12, 1999 (Pavlakis and Felber)

WO 98/17816 Lentiviral Vectors (Kingsman & Kingsman) (Oxford Biomedica Ltd)

WO 98/34640 (Shiver, J.W., Davies, M-E M., Freed, D.C., Liu, M.A. and Perry, H.C. - Merck & Co., Inc.)

WO 98/46083 Use of Lentiviral Vectors for Antigen Presentation in Dendritic Cells (Wong-Staal, Li; Kan-Mitchell) (Univ. of Cal.)

WO 99/04026 Lentiviral Vectors (Chen, Gasmi, Yee and Jolly) (Chiron)

WO 99/15641 Non-Primate Lentiviral Vectors and Packaging Systems (Poeschia, Looney and Wong-Staal) (Univ. of Cal.)

WO 99/30742 Therapeutic Use of Lentiviral Vectors (Naldini and Song)

WO 99/51754 Infectious Pseudotyped Lentiviral Vectors Lacking Matrix Protein and Uses Thereof (Goettlinger, Reil and Bukovsky) (Dana Farber Cancer Inst Inc)

- 41 -

PCT/US99/11082 Post-Transcriptional Regulatory Elements and
Uses Thereof (Pavlakis and Nappi), filed May 22, 1999, published as WO 99/61596
on December 2, 1999

Akkina, R.K., Walton, R.W., Chen, M.L., Li, Q-X, Planelles, V and
Chen, I.S.Y., "High-efficiency gene transfer into CD34⁺ cells with a human
immunodeficiency virus type 1-based retroviral vector pseudotyped with vesicular
stomatitis virus envelope glycoprotein G," *J. Virol.* 70:2581-2585 (1996)

Amado, R.G. & Chen, I.S.Y., "Lentiviral vectors—the promise of
gene therapy within reach?," *Science* 285:674-676 (July 1999)

Donahue, R.E., An, D.S., Wersto, R.P., Agricola, B.A., Metzger,
M.E. and Chen, I.S.Y., "Transplantation of immunoselected CD34⁺ cells transduced
with a EGFP-expressing lentiviral vector in non-human primates," *Blood* 92(suppl.
1):383b, Abstract #4648.5 (1998)

Fox, J.L., "Researchers wary of fear-based ban on lentivirus gene
therapy," *Nature Biotechnology* 16:407-408 (1998)

Goldman, M.J., Lee, P.S., Yang, J.S. & Wilson, J.M., "Lentiviral
vectors for gene therapy of cystic fibrosis," *Hum Gene Ther.* 8, 2261-2268 (1997)

Hartikka J, Sawdey M, Cornefert-Jensen F, Margalith M, Barnhart K,
Nolasco M, Vahlsing HL, Meek J, Marquet M, Hobart P, Norman J, and Manthorpe
M., "An improved plasmid DNA expression vector for direct injection into skeletal
muscle," *Hum Gene Ther.* 7:1205-17 (1996)

Kafri, T., Blomer, U., Peterson, D.A., Gage, F.H. & Verma, I.M.,
"Sustained expression of genes delivered directly into liver and muscle by lentiviral
vectors," *Nat Genet.* 17, 314-317 (1997)

- 42 -

Kafri, T., van Praag, H., Ouyang, L., Gage, F.G. and Verma, I.M., "A packaging cell line for lentivirus vectors," *J. Virol.* 73:576-584 (1999)

Kim, V.N., Mitrophanous, K., Kingsman, S.M., and Kingsman, A.J., "Minimal Requirement for a Lentivirus Vector Based on Human Immunodeficiency Virus Type 1", *J. Virol.* 72:811-816 (1998)

Klimatcheva, E., Rosenblatt, JD. and Planelles, V., "Lentiviral vectors and gene therapy," *Frontiers in Bioscience* 4:d481-496 (June 1999)

Miyoshi, H., Takahashi, M., Gage, F.H. & Verma, I.M., "Stable and efficient gene transfer into the retina using an HIV-based lentiviral vector," *Proc Natl Acad Sci USA*. 94: 10319-10323 (1997)

Miyoshi, H., Blomer,U., Takahashi, M., Gage, F.H., and Verma, I.M., "Development of self-inactivating lentivirus vector," , *J Virol.* 72:8150-8157 (1998)

Miyoshi, H., Smith, K.A., Mosier, D.E., Verma, I.M. and Torbett, B.E., "Transduction of human CD34⁺ cells that mediate long-term engraftment of NOD/SCID mice by HIV vectors," *Science* 283:682-686 (1999)

Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M. & Trono, D., "In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector," *Science*. 272, 263-267 (1996)

Naviaux, R.K., Costanzi, E., Haas, M. and Verma, I., "The pCL vector system: rapid production of helper-free, high-titer, recombinant retroviruses," *J. Virol.* 70:5701-5705 (1996)

Pavlakis, G.N., Schneider, R.; Song, S., Nasioulas, G., Zolotukhin, A., Felber, B.K., Trauger, R., Cox, J., and Manthorpe, M., "Use of simple Rev-

- 43 -

independent HIV-1 gag expression vectors in gene therapy and gene vaccine applications," *Natl Conf Hum Retroviruses Relat Infect (2nd)*, Jan 29-Feb 2 (1995); 91.

Poeschla, E.M., Wong-Staal, F. & Looney, D.J., "Efficient transduction of nondividing human cells by feline immunodeficiency virus lentiviral vectors," *Nature Med.* 4:354-357 (1998)

Qiu, J. T., R. Song, M. Dettenhofer, C. Tian, T. August, B. K. Felber, G. N. Pavlakis and X. F. Yu, "Evaluation of novel human immunodeficiency virus type 1 Gag DNA vaccines for protein expression in mammalian cells and induction of immune responses," *J Virol.* 73: 9145-52 (Nov. 1999)

Reynolds, P.N. and Curiel, D.T., "Viral vectors show promise in Colorado," *Nature Biotechnology* 16:422-423 (1998)

Schneider, R., Campbell, M., Nasioulas, G., Felber, B.K., and Pavlakis, G.N., Inactivation of the human immunodeficiency virus type 1 inhibitory elements allows Rev-independent expression of Gag and Gag/protease and particle formation, "*J. Virol.* 71:4892-4903 (1997)

Schwartz, S., M. Campbell, G. Nasioulas, J. Harrison, B. K. Felber and G. N. Pavlakis, "Mutational inactivation of an inhibitory sequence in human immunodeficiency virus type-1 results in Rev-independent gag expression," *J. Virol.* 66:7176-7182 (1992)

Shiver, J.W., Yasutomi, Y., Free, D.C., Davies, M.-E., Perry, H.C., Pavlakis, G.N., Letvin, N.L., and Liu, M.A., "DNA Vaccine-Mediated Cellular Immunity Against HIV-1 gag and env", presented at the Conference on Advances in AIDS Vaccine Development: 8th Annual Meeting of the National Cooperative Vaccine Development Groups for AIDS (NCVDGs) from February 11-15, 1996.

- 44 -

Soneoka, Y., Cannon, P.M., Ransdale, E.E., Griffiths, J.C., Romano, G., Kingsman, S.M. and Kingsman, A.J., "A transient three-plasmid expression system for the production of high titer retroviral vectors," *Nuc. Acids Res.* 23:628-633 (1995).

Srinivasakumar, N., Chazal, N., Helga-Maria, C., Prasad, S., Hammarskjöld, M.-L., and Rekosh, D., "The Effect of Viral Regulatory Protein Expression on Gene Delivery by Human Immunodeficiency Virus Type 1 Vectors Produced in Stable Packaging Cell Lines," *J. Virol.*, 71:5841-5848 (1997)

Sutton, R.E., Wu, H.T., Rigg, R., Bohnlein, E. & Brown, P.O., "Human immunodeficiency virus type 1 vectors efficiently transduce human hematopoietic stem cells," *J. Virol.* 72, 5781-5788 (1998)

Tabernero, C., A. S. Zolotukhin, J. Bear, R. Schneider, G. Karsenty and B. K. Felber, "Identification of an RNA sequence within an intracisternal-A particle element able to replace Rev-mediated posttranscriptional regulation of human immunodeficiency virus type 1," *J Virol.* 71:95-101 (1997). . (see also my email message)

Takahashi, M.; Miyoshi, H.; Verma, I.M.; Gage, F.H., "Rescue from photoreceptor degeneration in the rd mouse by human immunodeficiency virus vector-mediated gene transfer," *J. Virol.* 73: 7812-7816 (Sept. 1999)

Uchida, N., Sutton, R.E., Friera, A.M., He, D., Reitsma, M.J., Chang, W.C., Veres, G., Scollay, R. & Weissman, I.L., "HIV, but not murine leukemia virus, vectors mediate high efficiency gene transfer into freshly isolated G0/G1 human hematopoietic stem cells," *Proc. Natl Acad. Sci. U.S.A.* 95, 11939-11944 (1998)

Valentin, A., W. Lu, M. Rosati, R. Schneider, J. Albert, A. Karlsson and G. N. Pavlakis. "Dual effect of interleukin 4 on HIV-1 expression: Implications

- 45 -

for viral phenotypic switch and disease progression," *Proc. Natl Acad. Sci. U.S.A.* 95: 8886-91 (1998)

White, S.M., Renda, M, Nam, N-Y, Klimatcheva, E., Hu, Y, Fisk, J, Halterman, M, Rimel, B.J., Federoff, H, Pandya, S., Rosenblatt, J.D. and Planelles, V, "Lentivirus vectors using human and simian immunodeficiency virus elements," *J. Virol.* 73:2832-2840 (Apr. 1999)

Wolff, J.A. and Trubetskoy, V.S., "The Cambrian period of nonviral gene delivery," *Nature Biotechnology* 16:421-422 (1998)

Zolotukhin, J., Valentin, A., Pavlakis, G. N. and Felber, B. K. "Continuous propagation of RRE(-)and Rev(-)RRE(-) human immunodeficiency virus type 1 molecular clones containing a *cis*-acting element of Simian retrovirus type 1 in human peripheral blood lymphocytes," *J. Virol.* 68:7944-7952 (1994)

Zufferey, R., Nagy, D., Mandel, R.J., Naldini, L. and Trono, D., "Multiply Attenuated Lentiviral Vector Achieves Efficient Gene-Delivery In Vivo", *Nature Biotechnology* 15:871-875 (1997)

Zufferey, R., Dull, T., Mandel, R.J., Bukovsky, A., Quiroz, D., Naldini, L. & Trono, D., "Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery," *J. Virol.* 72:9873-9880 (1998)

- 46 -

Those skilled in the art will recognize that any gene encoding a mRNA containing an inhibitory/instability sequence or sequences can be modified in accordance with the exemplified methods of this invention or their functional equivalents.

Modifications of the above described modes for carrying out the invention that are obvious to those of skill in the fields of genetic engineering, virology, immunology, medicine, and related fields are intended to be within the scope of the following claims.

Every reference cited hereinbefore throughout the application is hereby incorporated by reference in its entirety.

WHAT IS CLAIMED IS:

1. A nucleic acid construct comprising a HIV-1 *gag/pol* gene having the coding sequence of the *gag/pol* gene set forth in Figure 1.
2. A nucleic acid construct comprising a HIV-1 *pol* gene having the coding sequence of the *pol* gene set forth in Figure 2.
3. A nucleic acid construct comprising a SIV-1 *gag* gene having the coding sequence of the *gag* gene set forth in Figure 3.
4. A nucleic acid construct comprising an HIV or SIV 5' LTR, a packaging signal, a *gag/pol* gene comprising the sequence set forth in Figure 1, a 5' splice site, a 3' splice site, an *env* gene, a *tat* gene, a functional RNA transport element and a 3' HIV or SIV LTR, said nucleic acid construct being able to produce functional Gag, Pol and Env virion components.
5. A vector comprising the nucleic acid construct of Claim 1, 2, 3 or 4.
6. A transformed host cell comprising the nucleic acid construct of Claim 1, 2, 3 or 4.
7. A transformed host cell of Claim 6 wherein said cell is a eukaryote.
8. The host cell of Claim 7 wherein said cell is a human cell.
9. A transformed host cell of Claim 6 wherein said cell is a prokaryote.
10. The host cell of Claim 9 wherein said cell is E. coli.
11. A pharmaceutical composition comprising the nucleic acid construct of Claim 1, 2, 3 or 4 and a pharmaceutically acceptable carrier.
12. A method for inducing antibodies in a mammal comprising administering to a mammal a composition of claim 11, wherein said nucleic acid construct is present in an amount which is effective to induce said antibodies in said mammal.
13. A method for inducing cytotoxic T lymphocytes in a mammal comprising administering to a mammal a composition of claim 11, wherein said

- 48 -

nucleic acid construct is present in an amount which is effective to induce cytotoxic T lymphocytes in said mammal.

14. A vaccine composition for inducing immunity in a mammal against HIV infection comprising a pharmaceutically acceptable carrier and further comprising a therapeutically effective amount of a nucleic acid construct of Claim 1 capable of producing HIV Gag and Pol proteins in the absence of HIV Rev regulatory protein in a cell in vivo.

15. A vaccine composition for inducing immunity in a mammal against HIV infection comprising a pharmaceutically acceptable carrier and further comprising a therapeutically effective amount of a nucleic acid construct of Claim 2 capable of producing HIV Pol protein in the absence of HIV Rev regulatory protein in a cell in vivo.

16. A vaccine composition according to claim 14 wherein said mammal is a human.

17. A vaccine composition according to claim 15 wherein said mammal is a human.

18. A method for inducing immunity against HIV infection in a mammal which comprises administering to a mammal a therapeutically effective amount of a vaccine composition according to claim 14.

19. A method for inducing immunity against HIV infection in a mammal which comprises administering to a mammal a therapeutically effective amount of a vaccine composition according to claim 15.

20. A method according to claim 18 wherein said mammal is a human.

21. A method according to claim 19 wherein said mammal is a human.

22. A lentiviral expression system comprising the following:

- (a) a packaging vector comprising a HIV-1 *gag/pol* gene having the nucleotide sequence set forth in Figure 1;
- (b) a transfer vector; and
- (c) an envelope encoding vector.

- 49 -

23. A transformed host cell comprising the lentiviral expression system of Claim 22.
24. A transformed host cell of Claim 23 wherein said cell is a eukaryote.
25. The host cell of Claim 24 wherein said cell is a human cell.
26. A process for making a lentiviral particle comprising expressing HIV Gag and HIV Pol in a host cell from a vector comprising the nucleotide sequences encoding HIV Gag and HIV Pol set forth in Figure 1 in the presence of a gene encoding an envelope protein.
27. A lentiviral expression system which is capable of functioning in the absence of Rev, Tat, and any viral RNA transport element comprising the following:
 - (a) a packaging vector comprising a HIV-1 *gag/pol* gene which has been mutated to eliminate inhibitory/instability regions;
 - (b) a transfer vector; and
 - (c) an envelope encoding vector.
28. A transformed host cell comprising the lentiviral expression system of Claim 27.
29. A transformed host cell of Claim 28 wherein said cell is a eukaryote.
30. The host cell of Claim 29 wherein said cell is a human cell.
31. A process for making a lentiviral particle in the absence of Rev, Tat, or any viral RNA transport element comprising expressing HIV Gag and HIV Pol in a host cell from a HIV-1 *gag/pol* gene which has been mutated to eliminate inhibitory/instability regions and expressing an Envelope protein from a envelope encoding gene whose expression is independent Rev, Tat, or any viral RNA transport element.
32. A nucleic acid construct comprising a SIV-1 *env* gene having the coding sequence of the *env* gene set forth in Figure 16.
33. A vector comprising the nucleic acid construct of claim 32.
34. A transformed host cell comprising the nucleic acid construct of claim 32.

- 50 -

35. A pharmaceutical composition comprising the nucleic acid construct of claim 32 and a pharmaceutically acceptable carrier.

1/43

FIGURE 1

ATGGGTGCGAGAGCGTCAGTATTAAAGCGGGGGAGAATTAGATCGATGGGAAAAAAT
TCGGTTAAGGCCAGGGGGAAAGAAGAAGTACAAGCTAAAGCACATCGTATGGCAA
GCAGGGAGCTAGAACGATTGCAGTTAACCTCTGGCTGTTAGAAACATCAGAAGGC
TGTAGACAAATACTGGGACAGCTACAACCATCCCTTCAGACAGGATCAGAGGAGCT
TCGATCACTATAACACAGTAGCAACCCTCTATTGTGTGCACCAGCGGATCGAGA
TCAAGGACACCAAGGAAGCTTAGACAAGATAGAGGAAGAGCAAAACAAGTCCAAG
AAGAAGGCCAGCAGGCAGCAGCTGACACAGGACACAGCAATCAGGTCAAGCAGC
TTACCCCTATAGTCAGAACATCCAGGGCAAATGGTACATCAGGCCATATCACCTA
GAACCTTAAATGCATGGTAAAAGTAGTAGAAGAGAAGGCTTCAGCCCAGAAGTG
ATACCCATGTTTCAGCATTATCAGAAGGAGCCACCCACAGGACCTGAACACGAT
GTTGAACACCGTGGGGGACATCAAGCAGCCATGCAAATGTTAAAAGAGACCATCA
ATGAGGAAGCTGCAGAATGGATAGAGTCAGTCAGTCAGGGCTATTGCA
CCAGGCCAGATGAGAGAACCAAGGGAAAGTGACATAGCAGGAACACTAGTACCC
TCAGGAACAAATAGGATGGATGACAATAATCCACCTATCCAGTAGGAGAGATCT
ACAAGAGGTGGATAATCCTGGATTGAACAAGATCGTGAGGATGTAGCCCTACC
AGCATTCTGGACATAAGACAAGGACCAAGGAACCCCTTAGAGACTATGTAGACCG
GTTCTATAAAACTCTAAGAGCTGAGCAAGCTTCACAGGAGGTAAAAATTGGATGA
CAGAAACCTTGGTCCAAATGCGAACCCAGATTGTAAGACCATCCTGAAGGCT
CTCGGCCAGCGGCTACACTAGAAGAAATGATGACAGCATGTCAGGGAGTAGGAGG
ACCCGGCCATAAGGCAAGAGTTGGCGAGGCGATGAGCCAGGTGACGAACCTGG
CGACCATAATGATGCAAGAGAGGCAACTCCGGAACCGCGGAAGATCGTCAAGTGC
TTCAATTGTGGCAAAGAAGGGCACACCGCCAGGAACCTGCCGGGGCCCCCGGAAGAA
GGGCTGTTGGAAATGTGGAAAGGAAGGACACCAAATGAAAGATTGTACTGAGAGAC

2/43

AGGCTAATTTTAGGAAAGATCTGGCCTTCCTACAAGGGAAGGCCAGGGATT
CTTCAGAGCAGACCAGAGCCAACAGCCCCACCAGAACAGAGAGCTTCAGGTCTGGGGT
AGAGACAACAACCTCCCCCTCAGAACGCAGGAGCCGATAGACAAGGAACGTATCCTT
TAACCTCCCTCAGATCACTCTTGGCAACGACCCCTCGTCACAGTAAGGATCGGGG
GGCAACTCAAGGAAGCGCTGCTCGATAACAGGAGCAGATGATAACAGTATTAGAAGAA
ATGAGTTGCCAGGAAGATGGAAACCAAAATGATAGGGGGATCGGGGCTTCAT
CAAGGTGAGGCAGTACGACCAGATACTCATAGAAATCTGTGGACATAAGCTATAG
GTACAGTATTAGTAGGACCTACACCTGTCAACATAATTGGAAGAAATCTGTTGACC
CAGATCGGCTGCACCTTGAACCTCCCCATCAGCCCTATTGAGACGGTGCCGTGAA
GTTGAAGCCGGGGATGGACGGCCCCAAGGTCAAGCAATGGCATTGACGAAAGAGA
AGATCAAGGCCTAGTCGAAATCTGTACAGAGATGGAGAACAGGAAGGAAAGATCAGC
AAGATCGGGCTGAGAACCCCTACAACACTCCAGTCTCGCAATCAAGAAGAAGGA
CACTACCAAGTGGAGAAAGCTGGTGGACTTCAGAGAGCTGAACAAGAGAACTCAGG
ACTTCTGGGAAGTTCAGCTGGCATCCCACATCCCGCTGGTTGAAGAAGAAGAAG
TCAGTGACAGTGCTGGATGTGGTGTGCCTACTTCTCCGTTCCCTGGACGAGGA
CTTCAGGAAGTACACTGCCTCACGATACTAGCATCAACAACGAGACACCAGGCA
TCCGCTACCAAGTACAACGTGCTGCCACAGGGATGGAAGGGATCACCAGCCATCTT
CAAAGCAGCATGACCAAGATCCTGGAGCCCTCCGCAAGCAAACCCAGACATCGT
GATCTATCAGTACATGGACGACCTCTACGTAGGAAGTGACCTGGAGATCGGGCAGC
ACAGGACCAAGATCGAGGAGCTGAGACAGCATCTGTTGAGGTGGGACTGACCACA
CCAGACAAGAACGCACCAAGAACCTCCCTCCTGTGGATGGCTACGAAC TGCA
TCCTGACAAGTGGACAGTGCAGCCATCGTGCCTGAGAACGGACAGCTGGACTG
TGAACGACATACAGAACGCTCGTGGCAAGTTGAACGGCAAGCCAGATCTACCCA
GGCATCAAAGTTAGGCAGCTGTGCAAGCTGCTCGAGGAACCAAGGCAGTGACAGA

3/43

AGTGATCCCAC TGACAGAGGAAGCAGAGCTAGAACTGGCAGAGAACCGAGAGATCC
TGAAGGAGCCAGTACATGGAGTGTACTACGACCCAAGCAAGGACCTGATGCCAGAG
ATCCAGAAGCAGGGCAAGGCCAATGGACCTACCAAATCTACCAGGAGGCCCTCAA
GAACCTGAAGACAGGCAGTACGCAAGGATGAGGGTGCCCACACCAACGATGTGA
AGCAGCTGACAGAGGCAGTGCAGAAGATCACCACAGAGAGCATCGTATCTGGGC
AAGACTCCCAAGTTCAAGCTGCCATACAGAAGGAGACATGGGAGACATGGTGGAC
CGAGTACTGGCAAGCCACCTGGATCCCTGAGTGGAGTCGTGAACACCCCTCCCT
TGGTGAACACTGTGGTATCAGCTGGAGAAGGAACCCATCGTGGGAGCAGAGACCTTC
TACGTGGATGGGCAGCCAACAGGGAGACCAAGCTGGCAAGGCAGGCTACGTGAC
CAACCGAGGACGACAGAAAGTGGTACCCCTGACTGACACCACCAACCAGAAAGACTG
AGCTGCAAGCCATCTACCTAGCTCTGCAAGACAGCGGACTGGAAGTGAACATCGT
ACAGACTCACAGTACGCAGTGGCATCATCCAAGCACAACCAGACCAATCCGAGTC
AGAGCTGGTGAACCAAGATCATCGAGCAGCTGATCAAGAAGGAGAAAGTGTACCTGG
CATGGGTACCAGCACACAAAGGAATTGGAGGAAATGAACAAGTAGATAAAATTAGTC
AGTGCTGGATCCGGAAAGGTGCTGTTCTGGACGGATCGATAAGGCCAAGATGA
ACATGAGAAGTACCACTCCAACCTGGCGCCTATGCCAGCGACTTCAACCTGCCAC
CTGTAGTAGCAAAAGAAATAGTAGCCAGCTGTGATAAAATGTCAGCTAAAAGGAGAA
GCCATGCATGGACAAGTAGACTGTAGTCCAGGAATATGGCAGCTGGACTGCACGCA
CCTGGAGGGAAAGGTGATCCTGGTAGCAGTTCATGTAGCCAGTGGATATATAGAAG
CAGAAGTTATCCCTGCTGAAACTGGCAGGAAACAGCATATTTCTTTAAAATTA
GCAGGAAGATGCCAGTAAAACAATACACACGGACAACGGAAAGCAACTTCACTGG
TGCTACGGTTAAGGCCGCTGTTGGTGGCGGGAAATCAAGCAGGAATTGGAATTC
CCTACAATCCCCAATCGCAAGGAGTCGTGGAGAGCATGAACAAGGAGCTGAAGAAG
ATCATCGGACAAGTGAGGGATCAGGCTGAGCACCTGAAGACAGCAGTGCAGATGGC

4/43

AGTGTTCATCCACAACTTCAAAAGAAAAGGGGGGATTGGGGGTACAGTGCAGGG
AAAGGATCGTGGACATCATGCCACCGACATCCAAACCAAGGAGCTGCAGAACGAG
ATCACCAAGATCCAGAACCTCCGGGTGTACTACCGCGACAGCCGCAACCCACTGTG
GAAGGGACCAGCAAAGCTCCTCTGGAAAGGGAGAGGGGCAGTGGTGATCCAGGACA
ACAGTGACATCAAAGTGGTGCCAAGGCGCAAGGCCAAGATCATCCGCGACTATGGA
AAACAGATGGCAGGTGATGATTGTGTGGCAAGTAGACAGGATGAGGATTAGAACCT
GGAAGAGCCTGGTGAAGCACCATATG (SEQUENCE ID NO:1)

5/43

FIGURE 2

>wildtype	TGTACAGAGA TGGAAAAGGA AGGGAAAATT TCAAAAATTG
>mutated	TGTACAGAGA TGGAGAAGGA AGGGAAAGATC AGCAAGATCG
#1*.....*
>wildtype	GGCCTGAAAA TCCATACAAT ACTCCAGTAT TTGCCATAAA
>mutated	GGCCTGAGAA CCCCTACAAC ACTCCAGTCT TCGCAATCAA
#41*.....*
>wildtype	GAAAAAAAGAC AGTACTAAAT GGAGAAAATT AGTAGATTTC
>mutated	GAAGAAGGAC AGTACCAAGT GGAGAAAGCT GGTGGACTTC
#81*.....**.....*
>wildtype	AGAGAACTTA ATAAGAGAAC TCAAGACTTC TGGGAAGTTTC
>mutated	AGAGAGCTGA ACAAGAGAAC TCAGGACTTC TGGGAAGTTTC
#121*.....*
>wildtype	AATTAGGAAT ACCACATCCC GCAGGGTTAA AAAAGAAAAAA
>mutated	AGCTGGGCAT CCCACATCCC GCTGGGTGA AGAAGAAGAA
#161*.....*
>wildtype	ATCAGTAACA GTACTGGATG TGGGTGATGC ATATTTTCA
>mutated	GTCAGTGACA GTGCTGGATG TGGGTGATGC CTACTTCTCC
#201*.....*
>wildtype	GTTCCCTTAG ATGAAGACTT CAGGAATAT ACTGCATTAA
>mutated	GTTCCCTTGG ACGAGGACTT CAGGAAGTAC ACTGCCTTCA
#241*.....*
>wildtype	CCATACCTAG TATAAACAAAT GAGACACCAAG GGATTAGATA
>mutated	CGATACCTAG CATCAACAAAC GAGACACCAAG GCATCCGCTA
#281*.....*
>wildtype	TCAGTACAAT GTGCTTCCAC AGGGATGGAA AGGATCACCA
>mutated	CCAGTACAAC GTGCTGCCAC AGGGATGGAA GGGATCACCA
#321*.....*
>wildtype	GCAATATTCC AAAGTAGCAT GACAAAAATC TTAGAGCCTT
>mutated	GCCATCTTTC AAAGCAGCAT GACCAAGATC CTGGAGCCCT
#361*.....*
>wildtype	TTAGAAAACA AAATCCAGAC ATAGTTATCT ATCAAATACAT
>mutated	TCCGCAAGCA AAACCCAGAC ATCGTGATCT ATCAGTACAT
#401*.....*

6/43

>wildtype	GGATGATTG TATGTAGGAT CTGACTTAGA AATAGGGCAG
>mutated	GGACGACCTC TACGTAGGAA GTGACCTGGA GATCGGGCAG
#441 * * * * * * * * *
>wildtype	CATAGAACAA AAATAGAGGA GCTGAGACAA CATCTGTTGA
>mutated	CACAGGACCA AGATCGAGGA GCTGAGACAG CATCTGTTGA
#481 * * * * * * *
>wildtype	GGTGGGGACT TACCACACCA GACAAAAAAC ATCAGAAAGA
>mutated	GGTGGGGACT GACCACACCA GACAAGAAGC ACCAGAAGGA
#521 * * * * * * *
>wildtype	ACCTCCATTC CTTTGGATGG GTTATGAAC CCATCCTGAT
>mutated	ACCTCCCTTC CTGTGGATGG GCTACGAAC GCATCCTGAC
#561 * * * * * * *
>wildtype	AAATGGACAG TACAGCCTAT AGTGCTGCCA GAAAAAGACA
>mutated	AAGTGGACAG TGCAAGCCCAT CGTGCTGCCT GAGAAGGACA
#601 * * * * * * *
>wildtype	GCTGGACTGT CAATGACATA CAGAAAGTTAG TGGGGAAATT
>mutated	GCTGGACTGT GAACGACATA CAGAAAGCTCG TGGGCAAGTT
#641 * * * * * * *
>wildtype	GAATTGGGCA AGTCAGATT ACCCAGGGAT TAAAGTAAGG
>mutated	GAACTGGGCA AGCCAGATCT ACCCAGGCAT CAAAGTTAGG
#681 * * * * * * *
>wildtype	CAATTATGTA AACTCCTTAG AGGAACCAAA GCACTAACAG
>mutated	CAGCTGTGCA AGCTGCTTCG AGGAACCAAG GCACTGACAG
#721 * * * * * * *
>wildtype	AAGTAATACC ACTAACAGAA GAAGCAGAGC TAGAACTGGC
>mutated	AAGTGATCCC ACTGACAGAG GAAGCAGAGC TAGAACTGGC
#761 * * * * * * *
>wildtype	AGAAAACAGA GAGATTCTAA AAGAACCAAGT ACATGGAGTG
>mutated	AGAGAACCGA GAGATCCTGA AGGAGCCAGT ACATGGAGTG
#801 * * * * * * *
>wildtype	TATTATGACC CATCAAAAGA CTTAATAGCA GAAATACAGA
>mutated	TACTACGACC CAAGCAAGGA CCTGATCGCA GAGATCCAGA
#841 * * *** * * * * * * *

7/43

>wildtype	AGCAGGGGCA AGGCCAATGG ACATATCAA TTTATCAAGA
>mutated	AGCAGGGGCA AGGCCAATGG ACCTACCAAA TCTACCAGGA
#881*.....*.....*
>wildtype	GCCATTAAA AATCTGAAAA CAGGAAAATA TGCAAGAATG
>mutated	GCCCTTCAAG AACCTGAAGA CAGGCAAGTA CGCAAGGATG
#921*.....*.....*
>wildtype	AGGGTGCCC ACACTAATGA TGTAACACAA TTAACAGAGG
>mutated	AGGGTGCCC ACACCAACGA TGTAAGGCAG CTGACAGAGG
#961*.....*.....*
>wildtype	CAGTGCAAAA AATAACCACA GAAAGCATAG TAATATGGGG
>mutated	CAGTGCAGAA GATCACCACA GAGAGCATCG TGATCTGGGG
#1001*.....*.....*
>wildtype	AAAGACTCCT AAATTTAAC TGCCCCATACA AAAGGAAACA
>mutated	CAAGACTCCC AAGTTCAAGC TGCCCCATACA GAAGGAGACA
#1041*.....*.....*
>wildtype	TGGGAAACAT GGTGGACAGA GTATTGGCAA GCCACCTGG
>mutated	TGGGAGACAT GGTGGACCGA GTACTGGCAA GCCACCTGG
#1081*.....*
>wildtype	TTCCTGAGTG GGAGTTTGT AATACCCCTC CTTTAGTGAA
>mutated	TCCCTGAGTG GGAGTTCTG AACACCCCTC CCTTGGTGAA
#1121*.....*.....*
>wildtype	ATTATGGTAC CAGTTAGAGA AAGAACCCAT AGTAGGAGCA
>mutated	ACTGTGGTAT CAGCTGGAGA AGGAACCCAT CGTGGGAGCA
#1161*.....*.....*
>wildtype	GAAACCTTCT ATGTAGATGG GGCAGCTAAC AGGGAGACTA
>mutated	GAGACCTTCT ACGTGGATGG GGCAGCCAAC AGGGAGACCA
#1201*.....*
>wildtype	AATTAGGAAA AGCAGGATAT GTTACTAATA GAGGAAGACA
>mutated	AGCTGGCAA GGCAGGCTAC GTGACCAACC GAGGACGACA
#1241*.....*.....**.....*
>wildtype	AAAAGTTGTC ACCCTAACTG ACACAAACAA TCAGAAGACT
>mutated	GAAAGTGGTG ACCCTGACTG ACACCAACCA CCAGAAGACT
#1281*.....*.....*

8/43

>wildtype	GAGTTACAAG CAATTTATCT AGCTTGCAG GATTGGGAT
>mutated	GAGCTGCAAG CCATCTACCT AGCTCTGCAA GACAGCGGAC
#1321 * * * * * * * * * * * *
>wildtype	TAGAAGTAAA CATACTAAC AACTCACAAT ATGCATTAGG
>mutated	TGGAAGTGAA CATCGTGACA GACTCACAGT ACGCACTGGG
#1361 * * * * * * * * * * * *
>wildtype	AATCATTCAA GCACAACCAG ATCAAAGTGA ATCAGAGTTA
>mutated	CATCATCCAA GCACAACCAG ACCAATCCGA GTCAGAGCTG
#1401 * * * * * * * * * * * *
>wildtype	GTCAATCAAA TAATAGAGCA GTTAATAAAAA AAGGAAAAGG
>mutated	GTGAACCAAGA TCATCGAGCA GCTGATCAAG AAGGAGAAAG
#1441 * * * * * * * * * * * *
>wildtype	TCTATCTGGC ATGGGTACCA GCACACAAAG GAATTGGAGG
>mutated	TGTACTGGC ATGGGTACCA GCACACAAAG GAATTGGAGG
#1481 * *
>wildtype	AAATGAACAA GTAGATAAT TAGTCAGTGC TGGATCAGG
>mutated	AAATGAACAA GTAGATAAT TAGTCAGTGC TGGGATCCGG
#1521 * *
>wildtype	AAAGTACTAT TTTAGATGG AATAGATAAG GCCCAAGATG
>mutated	AAGGTGCTGT TCCTGGACGG GATCGATAAG GCCCAAGATG
#1561 * * * * * * * * * *
>wildtype	AAACATGAGAA ATATCACAGT AATTGGAGAG CAATGGCTAG
>mutated	AAACATGAGAA GTACCACTCC AACTGGCGCG CTATGCCAG
#1601 * * * * * * * * * *
>wildtype	TGATTTAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA
>mutated	CGACTTCAAC CTGCCACCTG TAGTAGCAAA AGAAATAGTA
#1641 * * *
>wildtype	GCCAGCTGTG ATAAATGTCA GCTAAAAGGA GAAGCCATGC
>mutated	GCCACCTGTG ATAAATGTCA GCTAAAAGGA GAAGCCATGC
#1681
>wildtype	ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAACTAGA
>mutated	ATGGACAAGT AGACTGTAGT CCAGGAATAT GGCAGCTGGA
#1721 * *

9/43

>wildtype >mutated #1761	TTGTACACAT TTAGAAGGAA AAGTTATCCT GGTAGCAGTT CTGCACGCAC CTGGAGGGGA AGGTGATCCT GGTAGCAGTT * * * * *
>wildtype >mutated #1801	CATGTAGCCA GTGGATATAT AGAACGAGAA GTTATTCCAG CATGTAGCCA GTGGATATAT AGAACGAGAA GTTATCCCTG * *
>wildtype >mutated #1841	CAGAAACAGG GCAGGAAACA GCATATTTTC TTTTAAAATT CTGAAACTGG GCAGGAAACA GCATATTTTC TTTTAAAATT *
>wildtype >mutated #1881	AGCAGGAAGA TGGCCAGTAA AAACAATACA TACAGACAAT AGCAGGAAGA TGGCCAGTAA AAACAATACA CACGGACAAC * *
>wildtype >mutated #1921	GGCAGCAATT TCACCAAGTGC TACGGTTAAG GCCGCCTGTT GGAAGCAACT TCACGGGTGC TACGGTTAAG GCCGCCTGTT * * **
>wildtype >mutated #1961	GGTGGGCCGG AATCAAGCAG GAATTTGGAA TTCCCTACAA GGTGGGCCGG AATCAAGCAG GAATTTGGAA TTCCCTACAA
>wildtype >mutated #2001	TCCCCAAAGT CAAGGAGTAG TAGAATCTAT GAATAAAGAA TCCCCAATCG CAAGGAGTCG TGGAGAGCAT GAACAAGGAG *** * * *** * *
>wildtype >mutated #2041	TTAAAGAAAA TTATAGGACA GGTAAAGAGAT CAGGCTGAAC CTGAAGAAGA TCATCGGACA ATGAGGGAT CAGGCTGAGC * * + * * * * * *
>wildtype >mutated #2081	ATCTTAAGAC AGCAGTACAA ATGGCAGTAT TCATCCACAA ACCTGAAGAC AGCAGTGCAG ATGGCAGTGT TCATCCACAA * * * * *
>wildtype >mutated #2121	TTTTAAAAGA AAAGGGGGGA TTGGGGGTA CAGTGCAGGG CTTCAAAAGA AAAGGGGGGA TTGGGGGTA CAGTGCAGGG *
>wildtype >mutated #2161	GAAAGAATAG TAGACATAAT AGAACACAGAC ATACAAACTA GAAAGGATCG TGGACATCAT CGCCACCGAC ATCCAAACCA * * * * * * *

10/43

>wildtype	AAGAATTACA AAAACAAATT ACAAAAATTC AAAATTTCG
>mutated	AGGAGCTGCA GAAGCAGATC ACCAAGATCC AGAACTTCCG
#2201*.....*.....*.....*.....*.....*.....*
>wildtype	GGTTTATTAC AGGGACAGCA GAAATCCACT TTGGAAAGGA
>mutated	GGTGTACTAC CGCGACAGCC GCAACCCACT GTGGAAGGGA
#2241*.....*.....*.....*.....*.....*.....*
>wildtype	CCAGCAAAGC TCCTCTGGAA AGGTGAAGGG GCAGTAGTAA
>mutated	CCAGCAAAGC TCCTCTGGAA GGGAGAGGGG GCAGTGGTGA
#2281*.....*.....*.....*.....*
>wildtype	TACAAGATAA TAGTGACATA AAAGTAGTGC CAAGAAGAAA
>mutated	TCCAGGACAA CAGTGACATC AAAGTGGTGC CAAGGCGCAA
#2321*.....*.....*.....*.....*.....**..*
>wildtype	AGCAAAGATC ATTAGGGATT ATGGAAAACA GATGGCAGGT
>mutated	GGCCAAGATC ATCCGCGACT ATGGAAAACA GATGGCAGGT
#2361*.....*.....*.....*
>wildtype	GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAAC
>mutated	GATGATTGTG TGGCAAGTAG ACAGGATGAG GATTAGAAC
#2401*
>wildtype	TGGAAAAGTT TAGTAAAACA CCATATG
>mutated	TGGAAGAGCC TGGTGAAGCA CCATATG
#2441*.....*.....*.....*

11/43

FIGURE 3

ATGGCGTGAGAAACTCCGTCTTGTCAAGGAAGAAAGCAGATGAATTAG
AAAAAAATTAGGCTACGACCCAACGGAAAGAAAAAGTACATGTTGAAGC
ATGTAGTATGGGCAGCAAATGAATTAGATAGATTGGATTAGCAGAAAG
CCTGTTGGAGAACAAAGAAGGATGTCAAAAATACTTCGGTCTTAGCT
CCATTAGTGCCAACAGGCTCAGAAAATTAAAAAGCCTTATAACTG
TCTGCGTCATCTGGTGCATTACGCAGAAGAGAAAAGTCAAACACACTGA
GGAAGCAAAACAGATAGTGCAGAGACACCTAGTGGTGGAAACAGGAAC
CACCGAAACCAGCCGAAGACCTCTCGACCAACAGCACCCTAGCGGC
AGAGGAGGAAACTACCCAGTACAGCAGATCGGTGGCAACTACGTCCAC
CTGCCACTGTCCCCGAGAACCTGAACGCTGGGTCAAGCTGATCGAGG
AGAAGAAGTTGGAGCAGAAGTAGTGCCAGGATTCCAGGCAGTGTAG
AAGGTTGCACCCCTACGACATCAACCAAGATGCTGAACTCGTTGGAGA
CCATCAGGCAGCTATGCAGATCATCCGTGACATCATCAACGAGGAGGCT
GCAGATTGGACTTGCAGCACCCACAACCAGCTCCACAACAAGGACAA
CTTAGGGAGCCGTAGGATCAGACATCGCAGGAACCACCTCCTCAGTTG
ACGAACAGATCCAGTGGATGTACCGTCAGCAGAACCCGATCCCAGTAGG
CAACATCTACCGTCGATGGATCCAGCTGGGTCTGCAGAAATGCGTCCGT
ATGTACAACCCGACCAACATTCTAGATGTAACAAAGGGCAAAAGAG
CCATTTCAGAGCTATGTAGACAGGTTCTACAAAAGTTAAGAGCAGAAC
AGACAGATGCAGCAGTAAAGAATTGGATGACTCAAACACTGCTGATTCA
AAATGCTAACCCAGATTGCAAGCTAGTGCTGAAGGGGCTGGGTGTGAAT
CCCACCCCTAGAAGAAATGCTGACGGCTTGTCAAGGAGTAGGGGGCCG
GGACAGAAGGCTAGATTAAATGGCAGAAGGCCCTGAAAGAGGGGACCAAGAAAGC
CCAGTGCCAATCCCTTTGCAAGCAGCCCAACAGAGGGGACACTCTGCAAGGCAATG
CAATTAAAGTGTGGAATTGTGGAAAGAGGGACACTCTGCAAGGCAATG
CAGAGCCCCAAGAAGACAGGGATGCTGGAAATGTGGAAAATGGACCA
TGTTATGGCCAATGCCCAGACAGACAGGCGGGTTTTAGGCCTTGGT
CCATGGGGAAAGAAGCCCCGCAATTCCCCATGGCTCAAGTGCATCAGG
GGCTGATGCCACTGCTCCCCAGAGGACCCAGCTGTGGATCTGCTAAA
GAACTACATGCAGTTGGCAAGCAGCAGAGAGAAAAGCAGAGAGAAAAG
CAGAGAGAAGCCTTACAAGGAGGTGACAGAGGATTGCTGCACCTCAAT
TCTCTTTGGAGGAGACCAGTAG

12/43

FIGURE 4

SIV gag	-----
#1 ATGGGCGTGAGAAACTCCGTCTGTCAGGGAAGAAAGCAG
SIV gag	-----
#41 ATGAATTAGAAAAAATTAGGCTACGACCCAACGGAAAGAA
SIV gag	-----
#81 AAAGTACATGTTGAAGCATGTAGTATGGCAGCAAATGAA
SIV gag	-----
#121 TTAGATAGATTTGGATTAGCAGAAAGCCTGTTGGAGAAC
SIV gag	-----
#161 AAGAAGGGATGTCAAAAATACTTCGGTCTTAGCTCCATT
SIV gag	-----
#201 AGTGCCAACAGGCTCAGAAAATTAAAAAGCCTTATAAT
SIV gag	-----
#241 ACTGTCTGCGTCATCTGGTGCATTACCGCAGAAGAGAAAG
SIV gag	-----
SIVgagDX..	-----
#281 TGAAACACACTGAGGAAGCAAAACAGATAGTGCAGAGACA
SIV gag	-----A-A-----T-----A--A
SIVgagDX..	-----C-C-----C-----G-G
#321 CCTAGGGTGGAACAGGAACMACMGAAACYATGCCRAAR
SIV gag	--AAG-A-----
SIVgagDX..	--CTC-C-----
#361 ACMWSTMGACCAACAGCACCATCTAGCGGCAGAGGGAGGAA

13/43

SIV gag -T-----A--A--A-----T-----
 SIVgagDX..-C-----G--G--C-----C-----
 #401
 AYTACCCAGTACARCARATMGGTGGTAACTAYGTCCACCT

SIV gag ----T-AAG-----AT-A--T--C----A--AT--
 SIVgagDX..----C-GTC-----CC-G--C--T----C--GC--
 #441
 GCCAYTRWSCCGAGAACMYTRAAYGCYTGGGTMAARYTG

SIV gag --A-----A-----A--T-----
 SIVgagDX..-C-----G-----G-C-----
 #481
 ATMGAGGARAAGAARTTYGGAGCAGAAGTAGTGCCAGGAT

SIV gag -T-----T-----T-----T--
 SIVgagDX..-C-----C-----C-----C--
 #521
 TYCAGGCACTGTCAGAAGGTTGCACCCCTAYGACATYAA

SIV gag T-----T-A--T--T--G-----A-----
 SIVgagDX..C-----C-G--C--C--T-----G-----
 #561
 YCAGATGYTRAAYTGYGTKGGAGACCATCARGCGGCTATG

SIV gag -----T--A-A--T--T--A-----
 SIVgagDX..---C--C-T--C--C--C-----
 #601
 CAGATYATCMGWGAYATYATMAACGAGGGAGGCTGCAGATT

SIV gag -----
 SIVgagDX..
 #641
 GGGACTTGCAGCACCCACAACCAGCTCCACAACAAGGACA

SIV gag -----T--T-----A--T-----
 SIVgagDX..-----C--C-----C--C--
 #681
 ACTTACGGAGCCGTCAGGATCAGAYATYGCAGGAACMACY

SIV gag AGT-----A--T--A-----A-A--A--
 SIVgagDX..TCC-----T--C--G-----C-T--G--
 #721
 WSYTCAGTWGAYGAACARATCCAGTGGATGTACMGWCARC

14/43

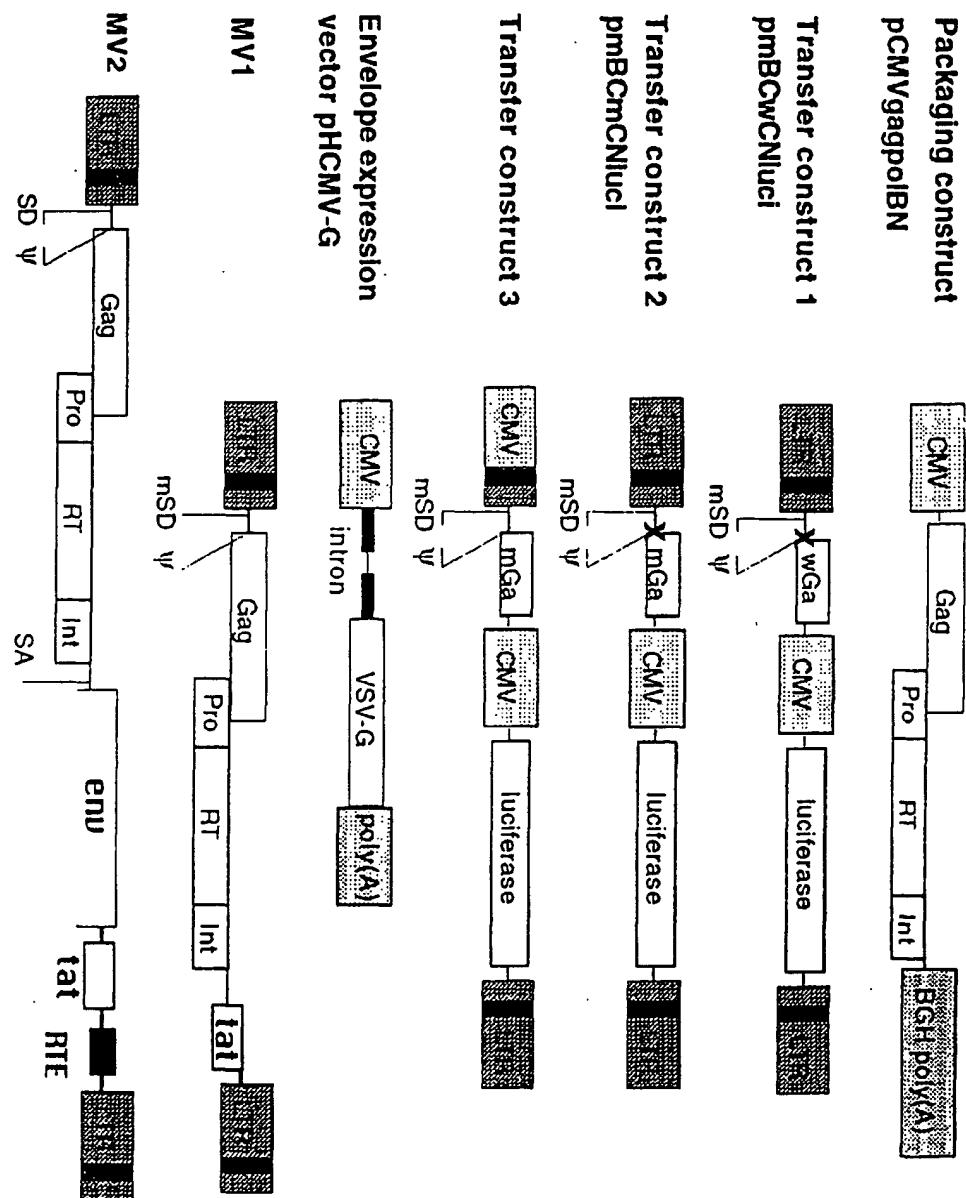
SIV gag	-----C-A-----T-A-GA-----
SIVgagDX..	-----G-C-----C-C-TC-----
#761
	AGAACCCSATMCCAGTAGGCAACATYACMGKMGATGGAT
SIV gag	---A-----GT---A-A-T-CA-A---T---A
SIVgagDX..	---G-----TC---G-G-C-TC-T---C---G
#801
	CCARCTGGGKYTGCARAARTGYGTYMGWATGTAYAACCR
SIV gag	--A-----
SIVgagDX..	--C-----
#841
	ACMAACATTCTAGATGTAAAACAAGGGCCAAAAGAGCCAT
SIV gag	-----
#881
	TTCAGAGCTATGTAGACAGGTTCTACAAAAGTTAACAGC
SIV gag	-----
#921
	AGAACAGACAGATGCAGCAGTAAAGAATTGGATGACTCAA
SIV gag	-----
#961
	ACACTGCTGATTCAAAATGCTAACCCAGATTGCAAGCTAG
SIV gag	-----
#1001
	TGCTGAAGGGCTGGGTGTGAATCCCACCCCTAGAAGAAAT
SIV gag	-----
#1041
	GCTGACGGCTTGTCAAGGAGTAGGGGGCCGGACAGAAG
SIV gag	-----
#1081
	GCTAGATTAATGGCAGAACGCCCTGAAAGAGGCCCTCGCAC
SIV gag	-----
#1121
	CAGTGCCAATCCCTTGCAGCAGCCAACAGAGGGGACC
SIV gag	-----
#1161
	AAGAAAGCCAATTAAAGTGTGGAATTGTGGAAAGAGGGGA

15/43

SIV gag #1201	----- CACTCTGCAAGGCAATGCAGAGCCCCAAGAACAGACAGGGAT
SIV gag #1241	----- GCTGGAAATGTGGAAAAATGGACCATTTATGGCCAAATG
SIV gag #1281	----- CCCAGACAGACAGGCAGGTTTTAGGCCTTGGTCCATGG
SIV gag #1321	----- GGAAAGAAGCCCCGCAATTCCCCATGGCTCAAGTGCATC
SIV gag #1361	----- AGGGGCTGATGCCAACTGCTCCCCAGAGGACCCAGCTGT
SIV gag #1401	----- GGATCTGCTAAAGAACTACATGCAGTTGGCAAGCAGCAG
SIV gag #1441	----- AGAGAAAAGCAGAGAGAAAGCAGAGAGAACGCCTTACAAGG
SIV gag #1481	----- AGGTGACAGAGGATTGCTGCACCTCAATTCTCTTTGG
SIV gag #1521	----- AGGAGACCAGTAG

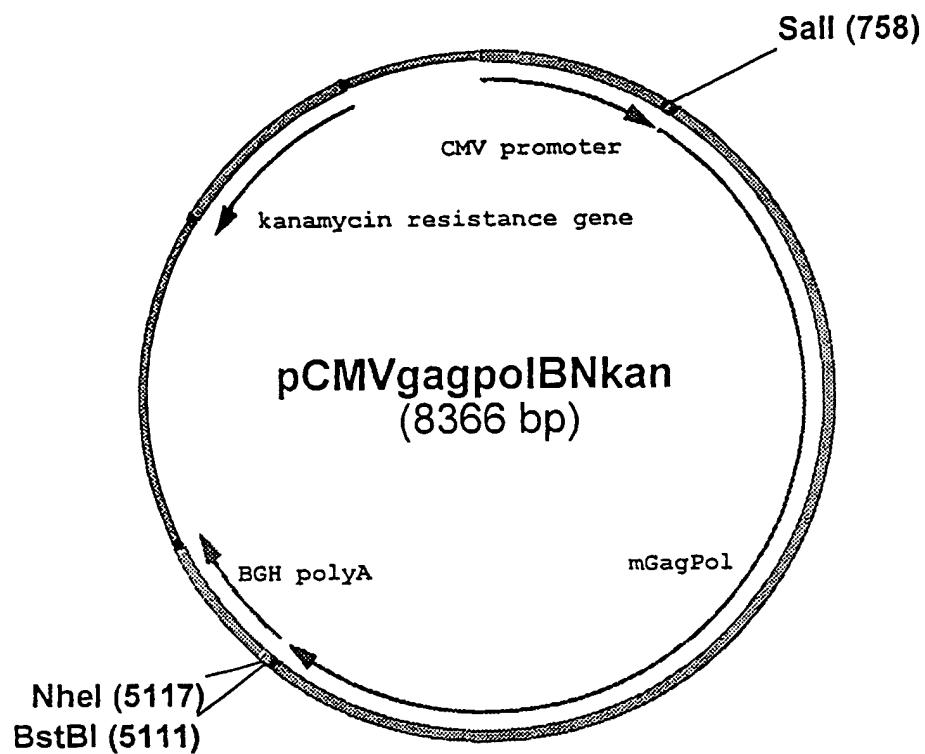
16/43

FIGURE 5



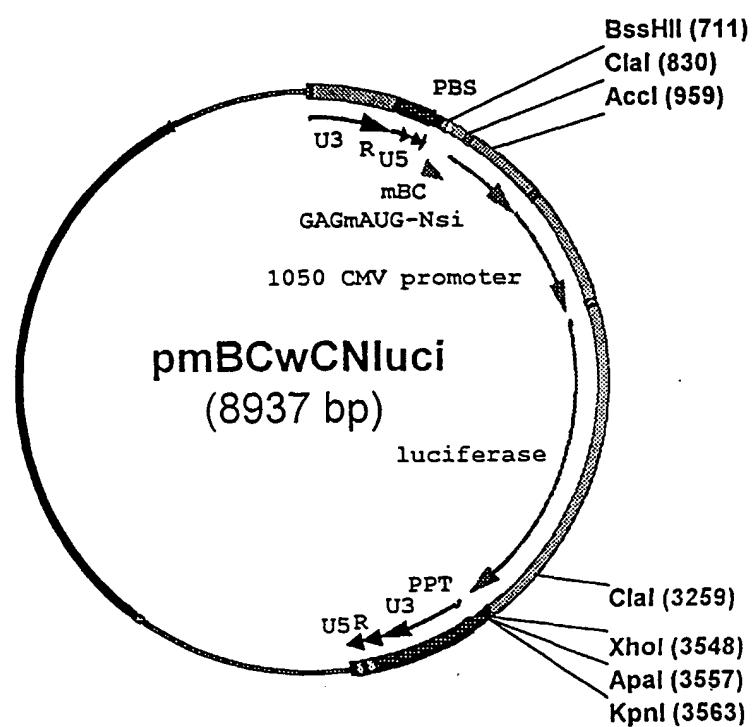
17/43

FIGURE 6



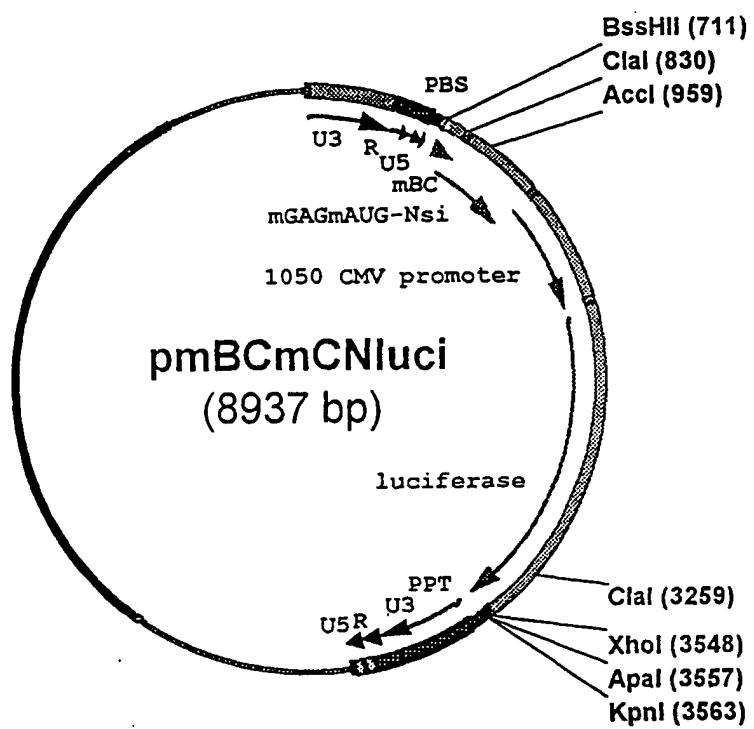
18/43

FIGURE 7



19/43

FIGURE 8



20/43

FIGURE 9

1 CCTGGCCATT GCATACGTTG TATCCATATC ATATAATGTA CATTATATT GGCTCATGTC CAACTTAC
 71 CCCATGTGA CATTGATTAT TGACTAGTTA TTAAATAGTAA TCAATTACGG GGTCAATTAGT TCATAGCCC
 141 TATATGGAGT TCCGGCGTAC ATAACCTACG GTAAATGCC CGCCTGGGTG ACCGGCCAC GACCCCGGCC
 211 CATTGACGTC AATAATGACG TATGTCCCA TAGTAAACGCC AATAGGGACT TTCCATTGAC GTCAATGGGT
 281 GGAGTATTAA CGGTAAACTG CCCACTTGCG AGTACATCAA GTGTATCATA TGCCTAAGTAC GCCCCCTATT
 351 GGGTCAATG ACGGTAAATG GGGCGCTGG CATTATGCC AGTACATGAC CTATGGGAC TTTCCTACTT
 421 GGCAGTACAT CTACGTATTA GTCACTGCTA TTACCATGGT GATGGGGTTT TGCCTAATACA TCAATGGCG
 491 TGGATACGGG TTTGACTCAC GGGGATTTCC AAGTCTCCAC CCCATTGAGG TCAATGGGAG TTTGTTTGG
 561 CACCAAAATC AACGGGACTT TCCAAATGT CGTAACRAT CGGGCCATT GACGCCAAATG GGGGGTAGGC
 631 GTGTACGGTG GGAGGTCTAT ATAAGCAGAG CTCTTTAGT GAACCGTCAG ATCCCTGGA GACGCCATCC

Sall (758)

701 ACGCTGTTT GACCTCCATA GAGAGAACCG GGACCGATCC AGCCTCCGGG GGCAGGGTC GACAGAGAGA →
 771 TGGGTGGAG AGCGTCAGTA TTAAGGGGG GAGAATTAGA TGGATGGGAA AAAATTGGT TAAGGCCAGG
 841 GGGAAAGAG AAGTACAAAGC TAAAGCACAT CGTATGGCA AGCAGGGAGC TAGAACGATT CGCAGTTAAT
 911 CCTGGCTGT TAGAACATC AGAAGGGTGT AGAACAAATAC TGGGACAGCT ACACCCATCC CTTCAGACAG
 981 GATCAGAGGA GCTTCGATCA CTATACAAAC CAGTACCAAC CCTCTATTGT GTCCACCGC GGATCGAGAT
 1051 CAGGGACACC AAGGGAGCTT TAGAACAGT AGAGGAAGAG CAAAACAAGT CCAAGAAGAA GCCCCAGCAG
 1121 GCAGCACCTG ACACAGGACA CAGCAATCAG GTCAGCCAA ATTACCCCTAT AGTCAGAAC ATCCAGGGC
 1191 AAATGGTACA TCAAGGCCATA TCACTTAGA CTTTAATGCC ATGGTAAAAT GTAGTAGAAG AGAAGGCTTT
 1261 CAGCCCCAAAG GTGATAACCA TGTTTCAGC ATTATCAGAA GGAGCCACCC CACAGGACCT GAACACGATG
 1331 TTGAACACCG TGGGGGACA TCAAGCAGCC ATGCAATGT TAAAGAGAC CATCAATGAG GAAGCTGCAG
 1401 AATGGGATAG AGTGCATCCA GTGCATGCCAG GGCCTATTGC ACCAGGCCAG ATGAGAGAAC CAGGGGAAG
 1471 TGACATAGCA GGAACCTACTA GTACCCCTCA GGAAACAAATA GGATGGATGA CAATAATCC ACCTATCCC
 1541 GTAGGAGAGA TCTACAGAG GTGGATAATC CTGGATTGTA ACAAGATCGT GAGGATGTAT AGCCCTACCA
 1611 GCATTCCTGA CATAAGACAA GGACCAAAGG AACCTTTAG AGACTATGTA GACCGGTTCT ATAAAACCT
 1681 AAGAGCTGAG CAAAGCTTCAC AGGAGGTAA AAATTGGATG ACAGAAACCT TGTGGTCCA AAATGCGAAC
 1751 CCAGATTGTA AGACCACTCT GAAGGCTTC GGCCTGGGG CTACACTAGA AGAACATGATG ACAGCATGTC
 1821 AGGGAGTAGG AGGACCCGGC CATAAGGCA GAGTTTGGC CGAGGGATG AGCCAGGTGA CGAACTCGGC
 1891 GACCTAAATG ATGCAGAGAG GCAACTTCGG GAACCAAGGGG AAGATCGTCA AGTGCCTCAA TTGTGGCAA

21/43

1961 GAAGGGCACA CGGCCAGGAA CTGCCGGGCC CCCCGGAAGA AGGGCTGTTG GAAATGTGGA AAGGAGGGAC
 2031 ACCAAATGAA AGATTTGACT GAGAGACAGG CTAATTTTTT AGGGAAAGTC TGGCCTTCCT ACAACGGAAAG
 2101 GCCAGGGAAAT TTCTTTCAGA GCAGACCCAGA GCCAACAGCC CCACCAAGAG AGAGCTTCAG GTCTGGGGTA
 2171 GAGACAACAA CTCCCCCTCA GAAGCAGGAG CCGATAGACA AGGAACGTGA TCCCTTAACCT TCCCTCAGAT
 2241 CACTCTTGG CAACGACCCC TCTTCACAGT AAGGATCGGG GGGCAACTCA AGGAACGGCT GCTCGATAAC
 2311 GGAGCAGATG ATACAGTATT AGAAGAAATG AGTTTGCAG GAAGATGGAA ACCAAAATG ATAGGGGGGA
 2381 TCGGGGGCTT CATCAAGGTG AGCCAGTACG ACCAGATACT CATAGAAATC TGTGGACATA AAGCTATAGG
 2451 TACAGTATTA GTAGGACCTA CACCTGTCAA CATAATTGGA AGAAATCTGT TGACCCAGAT CGGCTGCACC
 2521 TTGAACCTTC CCATCAGCCC TATTGAGACG GTGCCCGTGA AGTTGAAGCC GGGGATGGAC GGGCCCAAGG
 2591 TCAAGCAATG GCCATTGACG AAAGAGAAAGA TCAAGGCTT AGTCGAAATC TGTACAGAGA TGGAGAAGGA
 2661 AGGGAGATC AGCAAGATCG GCGCTTGAGAA CCCCTACAC ACTTCAGTCT TCGCAATCAA GAAGAAGGAC
 2731 AGTACCAAGT GGAGAAAGCT GGTGGACTTC AGAGAGCTGA ACAAGAGAAC TCAGGACTTC TGGGAAGTTTC
 2801 AGCTGGCAT CCCACATCCC GCTGGGTGAGA AGAAGAAGAA GTCACTGACCA GTGCTGGATG TGGGTGATGC
 2871 CTACTTCTCC GTTCCCTTGG ACAGGAGCTT CAGGAAGTAC ACTGCCCTCA CGATACCTAG CATCAACAAAC
 2941 GAGACCCAG GCATCCGCTA CCAGTACAAC GTGCTGCCAC AGGGATGGAA GGGATCACCA GCCATCTTTC
 3011 AAAGCAGCAT GACCAAGATC CTGGAGCCCT TCCGGAAAGCA AAACCCAGAC ATCGTGTATCT ATCACTACAT
 3081 GGACGACCTC TACGTAGGAA GTGACCTGGG GATGGGGAG CACAGGACCA AGATCGAGGA GCTGAGACAG
 3151 CATCTGTTGA GGTGGGGACT GACCACACCA GACAAGAAC ACCAGAGGA ACCTCCCTTC CTGTGGATGG
 3221 GCTACGAACG GCATCCTGAC AAGTGGACAG TGCAAGCCAT CGTGTGCTCCT GAGAAGGACA GCTGGACTGT
 3291 GAACGACATA CAGAAGCTCG TGGCAAGTT GAACCTGGCA AGCCAGATCT ACCCAGGCAT CAAAGTTAGG
 3361 CAGCTGTGCA AGCTGCTTCG AGGAACCAAG GCACGTGACAG AAGTGTATCCC ACTGACAGAG GAAGCAGAGC
 3431 TAGAACTGCC AGAGAACCGA GAGATCTGAA AGGAGCCAGT ACATGGAGTG TACTACGACC CAAGCAAGGA
 3501 CCTGATGCC GAGATCCAGA ACCAGGGGCA AGGCCTATGG ACCTACCAA TCTACCAGGA GCGCTTCAG
 3571 AACCTGAAGA CAGGCAAGTA CGCAACGGATG AGGGGTGCC ACACCAACGA TGTGAGCGAG CTGACAGAGG
 3641 CAGTGCAGAA GATCAACACAA GAGAGCATCG TGATCTGGGG CAAAGACTCCC AAGTCAAGC TGCCCATACA
 3711 GAAGGAGACA TGGGAGACAT GGTGGACCGA GTACTGGCAA GCCACCTGGA TCCCTGAGTG GGAGTTCTG
 3781 AACACCCCTC CCTTGGTGAAC ACTGTGGTAT CAGCTGGAGA AGGAACCCAT CGTGGGAGCA GAGACCTTCCT
 3851 ACGTGGATGG GCGAGCCAAAC AGGGAGACCA AGCTGGGCAA GCGAGGCTAC GTGACCAACC GAGGACGACA
 3921 GAAAGTGGTG ACCCTGACTG ACACCAACAA CCAGAAGACT GAGCTGCAAG CCATCTACCT AGCTCTGCAA
 3991 GACAGCGGAC TGGAACTGAA CATCGTGACCA GACTCACAGT ACGCACTGGG CATCATCCAA GCACAACCGAG

22/43

4061 ACCAATCCGA GTCAGAGCTG GTGAACCGA TCATCGAGCA GCTGTATCAAG AAGGAGAAAG TGTACCTGGC
 4131 ATGGGTACCA GCACACAAAG GAATTGGGG AAATGAAACAA GTAGATAAT TAGTCAGTGC TGGGATCCGG
 4201 AACGTGCTGT TCTGGACGG GATCGATAAG GCCCAAGATG AACATGAGRA GTACCACTCC AACTGGCGCG
 4271 CTATGCCAG CGACTTCAAC CTGCCACCTG TAGTAGCAA AGAAAATAGTA GCGAGCTGTG ATAAATGTCA
 4341 GCTAAAAGGA GAAGCCATGC ATGGACAAGT AGACTGTAGT CCAGGAATAT GCGAGCTGGA CTGCACGCAC
 4411 CTGGAGGGGA AGGTGATCTT GGTTAGCAGTT CTTGTAGCCA TTGGGATATAT AGAGGAGAA GTTATCCCTG
 4481 CTGAAACTGG GCAGGAAACA GCATATTTC TTTTAAATT AGCAGGAAGA TGGCCAGTAA AAACAAATACA
 4551 CACGGACAAC GGAAGCAACT TCACTGGTGC TACGGTTAAG GCCGCCTGTT GTGGGGGGGG AATCAAGGAG
 4621 GAATTGGAA TTCCCTACAA TCCCCAATCG CAAGGAGTCG TGAGAGCAT GAACAAGGAG CTGAAGAAGA
 4691 TCATCGGACA AGTGAGGGAT CAGGCTGAGC ACCTGAAGAC AGCAGTGCAG ATGGCAGTGT TCATCCACAA
 4761 CTTCAAAAGA AAAGGGGGGA TTGGGGGGTA CAGTGCAGGG GAAAGGATCG TGGACATCAT CGCCACCGAC
 4831 ATCCAAACCA AGGAGCTGCA GAAGCAGATC ACCAAGATCC AGAACTTCG GGTGTACTAC CGCGACAGCC
 4901 GCAACCCACT GTGGAAGGG A CCAACAAAGC TCCCTGGAA GGGAGAGGGG GCAGTGGTGA TCCAGGACAA
 4971 CAGTGACATC AAAGTGGTGC CAAGGCGCA A GCGCAAGTC ATCCCGACT ATGGAAAACA GATGGCAGGT
 5041 GATGATTTGTG TGGCAAGTAG ACAGGATGAG GATTAGAACCC TGGGAAGAGCC TGGTGAAGCA CCATATGGCG

**NheI (5117)****BstBI (5111)**

5111 TTGGAAGCTA GCCTCGGAGAT CCAGACCTGGC TGTGGCTTCT AGTGGCCAGC CATCTGTTCT TTGGCCCTCC
 5181 CCCGTGGCTT CCTTGACCTT GCAAGGTGGC ACTCCCCACTG TCTTTCCTTA ATAAAATGAG GAAATTGGCAT
 5251 CGCATTTGTCT GAGTAGGTGT CATTCTATTG TGCGGGGTGG GGTGGGGAG CACAGCAAGG GGGAGGATIG
 5321 GGAAGACAT ACCAGGGCATG CTGGGGATCC GGTGGGCTCT ATGGCTACTC AGGTGTGAA GAAATTGACCC
 5391 GGTTCCTCTT GGGCGAGAA GAGCCAGGCA CATCCCCCTTC TCTTGACAC ACCTGTCCA CGCCCGCTGGT
 5461 TCTTAACTTCC AGCCCCACTC ATAGGACACT CATAGCTCAG GAGGCTCTGG CCTTCATCC CACCCGCTAA
 5531 AGTACTTGGAA GGGGTCTCTC CCTGGCTCTG GAGCCGACCA AACAGACCT ACCTCCCGAG AGTGGGAAGA
 5601 AATTAAAGCA AGATAGGCTA TTAACTGGAG AGGGAGAGAA ATGGCTCCA ACATGTGAGG AAGTAATGAG
 5671 AGAAATCTTA GAATTTCCTTC CGCTTCCCTCG CTCACTGACT CGCTGGCGTC GGTGGTTCGG CTGGGGCGAG



5741 CGGTATCAGC TCACCTAAAG GCGGTAATAC GGTATCCAC AGAACAGGG GATAACGCAG GAAAGAACAT
 5811 GTGAGCAAAA GCGCACAAA AGGGCAGGA CGTAAAAAG GCGGGTTTCG TGGCGTTTTT CCATAGGCTC
 5881 CGCCCCCTTG ACGGACATCA CAAAATCGA CGCTCAAGTC AGAGGTGGCG AAACCCGACA GGACTATAAA
 5951 GATACCAAGC GTTCCCCCTT GGAAGCTCCC TCTGTGGCTC TCCCTGGTCCG ACCTGGCGC TTACCGGATA
 6021 CCTGTCCGCC TTTCCTCTT CGGGAAAGCTG GGGCTTCT CAAATGCTCAC GCTGTAGGTA TCTCAGTTG
 6091 GTGTAGGTG TTCGCTCCAA GCTGGGTGT GTGCACGAAC CCCCCGTTCA GCGCGACCGC TGGCGCTTAT
 6161 CGGGTAACTA TGTCTTGTAG TCCACCCCG TAAGACACGA CTATCGCCA CTGGCAGCAG CCACTGGTAA
 6231 CAGGATTAGC AGAGGGAGGT ATGTAGGGGG TGCTACAGAG TTCTTGAAGT GGTGGCTAA CTACGGCTAC

23/43

6301 ACTAGAAGGA CAGTATTGG TACTCTGGCT CTGCTGAAGC CAGTTACCTT CGGAAAAAGA GTGGTAGCT
 6371 CTTGATCCGG CAAACAAACC ACCGCTGGTA GCGGTGGTT TTGTTGTTGC AAGCAGCAGA TTACGCGCAG
 6441 AAAAAAGGA TCTCAAGGAAG ATCCTTGTAT CTTTCTACG GGCTCTGACG CTCACTGAA CGAAAACCTCA
 6511 CGTTAAGGA TTGTTGTAT GAGATTATCA AAAAGGATCT TCACCTAGAT CCTTTAAAT TAAAATGAA
 6581 GTTTAAATC AATCTAAAGT ATATATGAGT AAACCTGGTC TGACAGTTAC CAATGCTTAA TCAGTGAGGC
 6651 ACCTATCTCA GCGATCTGTC TATTCGCTTC ATCCATAGTT CCTGTGACTCC GGGGGGGGGG GCGCTGAGG
 6721 TCTGGCTCGT GAAGAAGGTG TTGCTGACTC ATACCAAGGCC TGAATGCC CATCATCCAG CCAGAAAGTG
 6791 AGGGACCCAC GGTGATGAG ACCTTGTGAG TAGGTGGACG ACTTGGTGTAT TTGAACTTT TCTTTGCCA
 6861 CGGAACGGTC TGGCTGTGCG GGAAGATGCG TGATCTGATC CTTCACATCA GCAAAAGTTC GATTTATTCA
 6931 ACAAAACCCG CGTCCCGTCG ACTCAGGGTA ATGCTCTGCC ATGTTTACAA CCAATTAAACC AATTCGATT
 7001 AGAAAACCTC ATCGAGCATC AAATGAAACT GCAATTATT CATATCAGGA TTATCAATAC CATATTTTG
 271 PhePheGlu AspLeuMetL euHsPheGlu LeuLysAsn MetAspProA snAspIleGlyTyrLysGln
 7071 AAAAACCGT TTCTGTAATG AAGGAGAAAA CTACCGAGG CAGTTCATA GGATGGCAAG ATCTGGTAT
 248 PheLeuArgL ysGlyLeuSe rProSer Phe GluGlyLeuC ysAsnTrpLe uileAlaLeu AspGlynTyrA
 7141 CGGTCTGGCA TTCCGACTCG TCCACATCA ATACAACCTA TTATTTCCC CTCGTCAAAA ATAAGGTTAT
 224 rAspAlaI eGlyValArg GlyValAspI leCysGlyII eLeuGly GluAspPhel IleLeuAsnAs
 7211 CAAGTGAGAA ATACCATGTA GTGACGACTG AACCGGTGA GAAATGCCAA AGCTTATGCA TTCTTTCCA
 201 pLeuSer Phe AspGlyHisT hrValValSe rAspProSer PheProLeuL euLysHisMe IGlLysTrp
 7281 GACTTGTICA ACAGGCCAGC CATTACGCTC GTCATCAAA TCACTGGCAT CAACCAAACCC GTTATTCA
 178 ValGlyuV alProTrpGly AsnArgGlu AspAspPheA spSerAlaAs pValLeuGly AsnAsnMetA
 7351 CGTGATTGG CCTGAGGAGG AGGAAATACG CGATGGCTGT TAAAGGACA ATTACAAACA GGATCGAAT
 154 rgSerGlnAlaGlnAlaLeu ArgPheValA rgAspSerAs nPheProCys AsnCysValP roleSerH
 7421 GCAACCGGGC CAGGAACACT GCCAGGCCAT CAACAATATT TTCACCTGAA TCAGGATATT CTCTAAATAC
 131 sLeuArgArg LeuPheValA IaLeuAlaAs pValLeuAsn GluGlySerA spProTyrGluGlyLeuVal
 7491 CTGGATGCT GTTITCCCCG GGATGGCAGT GGTGAGTAAC CTGCATCAT CAGGAGTAGC GATAAAATGC
 108 GluNpheAlaT hrLysGlyPr oileAlaThr ThrLeuLeuT rPAIaAspAs pProThrArg llePheHisL
 7561 TTGATGGTCG GAAGAGGCAT AAATCCGTC AGCCAGTTTA GTCTGACCAT CTCATCTGTA ACATCATGG
 84 ysIleThrPr oLeuProMet PheGluThrL euTrpAsnLeuArgValMet GluAspThrValAspAsnAl
 7631 CAACGCTACC TTGCTCATGT TTCAAGAACAA ACTCTGGCGC ATCGGGCTTC CCATACAATC GATAGATTGT
 61 aValSerGly LysGlyHisLysLeuPheLe uGlyLeuProAla AspProLysGlyTyrLeuArgTyrIleTh
 7701 CGCACCTGAT TGCCCGACAT TATCCGAGC CCATTTATAC CCATATAAT CAGCATCCAT GTGGAATTT
 38 AlaGlySerGlnGlyValAsnAspArgAla TrpLysTyrGlyTyrLeuAsn PAIaAspMet AsnSerAsnL
 7771 AAATCCGGCC TCGAGCAAGA CGTTCCCGT TGATATGTC CCATAACACC CCTTGTATTA CTGTTATGT
 14 euArgProArgSerCysThr GluArgGlnIleHisMet
 7841 AAGCAGACAG TTGTTGATGATA TATTTTATC TTGTGCAATG TAACATCAGA GATTTTGAGA
 7911 CACAACTGTT CTTTCCCCC CCCCCCATTA TTGAGCATT TATCAGGGTT ATTGCTCAT GAGCGGATAC
 7981 ATATTGAAAT GTATTTGAA AAATAAACAA ATAGGGTTTC CGGCACATT TCCCCGAAAAA GTGCCACCTG
 8051 ACGCTAAAGA AACCAATTATC ATCATGACAT TAACCTATAA AAATAGGGCT ATCACGAGGC CCTTTCGCTC
 8121 CGCGGGTTTC GGTGATGAGG GTGAAACCT CTGACACATG CAGCTCCGG AGACGGTCAC AGCTTGTCTG
 8191 TAAGGGATG CGGGAGGAG ACAAGCCGT CAGGGCGGT CAGGGGGTGT TGGGGGTGT CGGGGCTGGC
 8261 TTAACATATGC GGCATCAGAG CAGTTGTAC TGAGAGTCGA CCATATCCGG TGTGAATAC CGCACAGATG
 8331 CGTAAGGAGA AAATACCGCA TCAGATTGGC TATGG

24/43

FIGURE 10

1 TGGAAAGGCT AATTGGTCC CAAAAAAGAC AAGAGATCT TGATCTGTGG ATCTACCACA CACRAGGCTA
 71 CTTCCCTGAT TGGCAGAACT ACACACCAGG GCCAGGGATC AGATATCCAC TGACCTTGG ATGGTGCTTC
 141 AAGTTAGTAC CAGTTGAACC AGAGCAAGTA GAAGAGGCCA ATAAGGAGA CAAGACAGC TTGTTACACC
 211 CTATGAGCCA GCATGGGATG GAGGACCCGG AGGGAGAAGT ATTAGTGTGG AAGTTTGACA GCCTCCCTAGC
 281 ATTTCGTCAC ATGGCCCGAG AGCTGCATCC GGAGTACTAC AAAGACTGCT GACATCGACG TTTCCTACAG
 351 GGACTTTCCG CTGGGGACTT TCCAGGGGG TGTTGGCTGG GCGGGACTGG CGAGTGGCGA GCCTCCAGAT
 421 GCTACATATA AGCAGCTGCT TTTTGCCTGT ACTGGGTCTC TCTGGTTAGA CCAGATCTGA GCCTGGGAGC
 491 TCTCTGGCTA ACTAGGGAAC CCCTGCTTA AGCCTCAATA AAGCTTGCT TGAGTGTCTA AAGTAGTGTG
 561 TGCCCGTCTG TTGTGTGACT CTGGTAACTA GAGATCCCTC AGACCCCTTT AGTCAGTGTG GAAAATCTCT
 631 AGCAGTGGCG CCCGAAACAGG GACTTGAAG CGAAAGTAAA GCCAGAGGAG ATCTCTCGAC GCAGGACTCG
 →
BsHII (711)
 701 GCTTGCTGAA GCGGCGGCGGG caaaagggaga ggggggggggC ctggACggGa cggaaaaat ttttgcgttgc

ClaI (830)

771 ggaggtttaga aggagagagC TGGCTGGAG AGCTCTATA TTTAGGGGG GGGTTTATA TCGATGGAA
 →
 841 AAAATTCGGT TAAGGCAAG CGAAAGAAA AAAATAATAAT TAAAGATAT AGTATGGGCA ASCAGGGAGC

AccI (959)

911 TAGAACGATT CGGAGTTAAT CCTGGCTCT TAAAGACATC AGAGGGTGT AGACAAATAC TGGGACAGCT
 981 ACACCCATCC CTTCTGACAG GATCAGAGA ACTTGGATCA TTTATATATA CACTGGCAC CCTCTATTTG
 1051 GTGCATCATA GGATAGAGAT AAAAGACCC AAGGGAGCT TAGACAAGAT AGAGGAAGAG CAAACAAAA
 1121 GTAAGAAAAA AGCACAGCAA GCACGGCTG ACACAGGACA CAGCACTAG GTCAGCCAA ATTACCCAT
 1191 AGTCAGAAC ATCCAGGGGC AAATGGTACA TCAGGCCATA TCACTAGAA CTTTAAACGA TAAGCTTGGG
 1261 AGTTCGGGT TACATAACTT AGGTTAAATG GCGGGCTGG CTGACCGGCC AACGACCCCCC GCCCATTTGAC
 1331 GTCATAATG ACGTATGTTG CCATAGTACG GCAATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT
 1401 TTACCGTAA CTGCCACTT GGCAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA
 1471 ATGACGGTAA ATGGCGCGCC TGCCATTATG CCCAGTACAT GACCTTATGG GACTTTCTTA CTTGGCAGTA
 1541 CATCTACGTA TTAGTCATCG CTATTACCAT GGTGATGGGG TTTGGCAGT ACATCAATGG GCGTGGATAG
 1611 CGGTTGACT CACGGGGATT TCAAGTCTC CACCCCATGG ACGTCAATGG GAGTTTGTCTT TGGCACCAAA
 1681 ATCAACGGGA CTTTCCAAA TGTCGTAACT ACTCGCCCC ATTGAGGAA ATGGCGGTG CCGGTGTACG
 1751 GTGGGAGGTC TATATAAGCA GAGCTGTCTT AGTGAACCGT CAGATGCCCT GGAGACGCCA TCCACGGCTGT

25/43

1821 TTTGACCTCC ATAGAAGACA CGGACTCTAG AGgatecATC TAATTAAGCT TGGCATTCGG CTACTGTGG
 1891 TAAAATGGAA GACGCCRAAA ACATAAAGAA AGGCCCGGGG CCATTTTATC CTCTAGAGGA TGGAAACGGT
 1961 GGAGAGCAAC TGCATAAGGC TATGAAGAGA TAGGCCCTGG TTCTCTGGAAC AATTGCTTT ACAGATGCAC
 2031 ATATCGAGGT GAACATCAGG TACGCCGAAT AGCTCGAAT GTCCCGTGGG TTGGCAGAAG CTATGAAACG
 2101 ATATGGGCTG AATACAAATC ACAGAATCGT CGTATGAGT GAAAACCTTC TTCAATTCCT TATGCCGGTG
 2171 TTGGGGCGGT TTTTATGGG AGTTGAGTT GGCCCCGGG AGGAGATTTA TAATGAACT GATTTGCTCA
 2241 ACAGTATGAA CATTGCGAG CCTACCGTAG TGTTTGTTC CAAACAGGGG TTGCAAAAAA TTTTGAACG
 2311 GCAAAAAAAA TTACCAATAA TCCAGAAAT TATTATCGT GATTGATAAA CGGATTACCA GGGATTTCAG
 2381 TGATGTACA CGTTCGTAC ATCTCATCTA CCTCCCCGGT TTAATGATA CGATTTGTA CGAGACTCT
 2451 TTGATCGTCA CAAACATT GCAGTGTAA TGAATTCTTC TGGATCTACT GGGTACCTA AGGGTGTGGC
 2521 CCTTCGGCAT AGAACTGCT CGGTCAGATT CTGGCATGCC AGAGATCTA TTITTCGCA TCAATCAATT
 2591 CGGGTACTG CGATTTTAAG TGTTGTTCCA TTCCATACG GTTTGGAAAT GTTTACTACA CTGGATATT
 2661 TGATATGTGG ATTTCGAGTC CTCTTAATGT ATAGATTGA AGAAGGCTG TTTTACGAT CCTTCAGGA
 2731 TTACAAATT CAAACTGGT TGCTAATACC AACCTTATT TCATTTTGG CCAAAAGCAC TCTGATTGAC
 2801 AAATACGATT TATCTAATT ACACGAATT GCTTGGGGG GGGACCTCTT TTGGAAGAA GTGGGGGAAG
 2871 CGGTGCAAAC AGCTTCCAT CTTCGAGGGG TACGAGCTT ATATGGCTC ACTGAGACTA CATAGCTTAT
 2941 TCTGATTACA CGCGAGGGGG ATGATAAACC GGGCGGGTC GTAAAGTTG TTCTTTTTTG TGAACGGAAG
 3011 GTTGTTGGATC TGGATACCCG GAAACGCTG GCGGTTAATC AGAGAGGGGA ATTATGTGTC AGAGGACCTA
 3081 TGATTATTC CGGTATGTA AACATCCG AAGCGACCA CGCGTTGATT GACAGGGATG GATGGCTACA
 3151 TTCTGGAGAC ATAGCTTACT CGGAGGAAGA CGACACATTG TTCTAGTTG ACCGCTTGAA CTCTTTAATT

ClaI (3259)

3221 AAATACAAAG GATATCAGGT GGGCCCCGGT GAATTGAAAT CGATATTGTT ACAACACCC AACATCTTCG
 3291 ACGGGGGGGT CGCAGGTCTT CCCCAGGATG AGCCCGGTGA ACTTCCCGCC GCGGTGTTG TTTGGAGCA
 3361 CGGAAAGACG ATGACGGAAA AAGAGATCTG GGATTACCTC GCCAGTCAG TAACAACCCG GAAAAGTTG
 3431 CGCGGAGGAG TTGTGTTGT GGACGAAGTA CGCGAGGTC TTACCGAAA ACTCGACGCA AGAAAATCA

Apal (3557)**Xhol (3548) KpnI (3563)**

3501 GAGAGATCTT CATAAAGCCC AAGAAGGGCG GAAAGTCCAA ATTGTAACTC GAGGGGGGC CGGGTACCTT
 3571 TAAGACCAAT GACTTACAG CGAGCTGTAG ATCTTAGCCA CTTTTTAAAGA GAAAAGGGGG GACTGGAAGG
 3641 GCTTAATTAC TCCCAAGAA GACAAGTAT CCTTGATCTG TGGATCTACC ACACACAGG CTACTTCCT

3711 GATTGGCAGA ACTACACACC AGGCCAGGG GTCAAGATTC CACTGACCTT TGGATGGTGC TACAAGCTAG
 3781 TACCA GTTGA GCCAGATAAG GTAGAAAGAGG CCAATAAAGG AGAGAACACC AGCTTGTAC ACCCTGTGAG
 3851 CCTGCATGGA ATGGATGACC CTGAGAGAGA AGTGTAGAG TGAGGTTTG ACAGCCCCCT AGCATTTCAT
 3921 CACGTGGCCC GAGAGCTGCA TCCGGAGTAC TTCAAGAACT GCTGACATCG AGCTTGTAC AAGGGACTTT
 3991 CGCGTGGGA CTTTCCAGGG AGGGTGGCC TGGGGGGAC TGGGGASTGG CGAGCCCTCA GATGCTGGCAT
 4061 ATAAGCAGCT GCTTTTGCC TGTACTGGT CTCCTGGT AGACCAGATC TGAGGCTGGG AGCTCTCTGG
 4131 CTAACTAGGG AACCCACTGC TTAGGCCTCA ATAAAGCTTG CCTTGAGTGC TTCAAGTAGT GTGTGCCCGT
 4201 CTGTTGTG TG ACTCTGGTAA CTAGAGATCC CTCAGACCCCT TTTAGTCAGT GTGGAAAATC TCTAGCACCC
 4271 CCCAGGAGGT AGAGGTGCA GTGAGGCCAG ATCGGCCAC TGCATTCAG CCTGGCAAG AAAACAAGAC
 4341 TGTCTAAAT AATAATAATA AGTAAAGGGT ATTAATAATA TTATACATG GAGGTCTATAA AAATATATAT
 4411 ATTTGGCTG GGCGCAGTGG CTCACACCTG CGCCCGGCC TTTGGGAGGC CGAGGAGGT GGATCACCTG
 4481 AGTTGGGAG TICCAAGACCA GCTGACCAA CATGGAGAAA CCCCTTCTCT GTGTTATTTT AGTAGATTTT
 4551 ATTTTATGTC TATTTTATTAC ACAGCTTATT CTGAGAAACT GAATCTGTT TTCTCTCTACT CTGATACAC
 4621 AAGAATCATC AGCACAGGG AAGACTTCTG TGATCAAATG TGTTGGAGA GGGAGTTTTT CACCAAGCACA
 4691 TGAGCAGTCA GTTCTGCCGC AGACTGGCG GGTGCTCTTC GGTTCAGTTC CAACACCGCC TGCTGGAGA
 4761 GAGGTCAAGAC CACACGGTGA GGCGTCACTC CCAAGAACAT AACACCCAA GACATAAAACA CCAACAGGT
 4831 CCACCCCGCC TGTCTCCAG GCAGAGCGGA TTCAACAAAGA CGGAATTAG GATAGAGAAA GAGTAAGTCA
 4901 CAACAGCGG CGCTGCGGG AGAACGGAGT TTATATGCA CTCAATCAG TCTCCCAAG CATTGGGGA
 4971 TCAAGCTTTT TAAGGATAAC TTAGTGTGA GGGGGCAGT GACTTGGAGA TGAACCGTA GGGAGTCGAA
 5041 GGTGTCCTTT TCGGCGAGT CAGTCTCTGG GTGGGGGCCA CAAGATEGGA TGAGGCTGGT TATCAATCCG
 5111 CGGGTCCAG CTTGTCATG GACTGCAAGG TCTGCAAAAT ATCTCAAGCA CTGATMGTAC TTAGGTTTTA
 5181 CTAATGTAT GTTACCCAG GAACAATTGG GGGAGGTCA GAATCTTGTA GCCTGTAGCT GCATGACTCC
 5251 TAAACCTAA TTCTCTTTT TTCTCTTTT TTCTCTTTT GAGACAGGGT CTCACTCTGT CACCTAGGT
 5321 GGAGTCCAGT GGTCAATCA CAGCTCACTG CAGCCCCCTAG AGGGCCGCC ACCGGCTGG AGCTCCAATT
 5391 CGCCCTATAG TGAGTCGTAT TACAATTCAC TGGCGCTGGT TTACAACTGT CGTGTACTGGG AAAACCCCTGG
 5461 CGTTACCAA CTTAATCGCC TTGCAAGCA TCCCTCTGG CCAAGCTGGC GTAAATAGCGA AGAGGCCCGC
 5531 ACCGATGCC CTTCCCAAAGA TTGCGCAGC CTGAATGGCG AATGGCGCA AATITGAAAC GTAAATATTT
 5601 TGTTAAATT CGCGTAAAT TTGTTTAAA TCAGCTCATI TTAAACCAA TAGGGCGAAA TCGGAAAT
 5671 CCCTTATAAA TCAAAAGAT AGACCGAGAT AGGGTGTAGT GTGTTCCAG TTGGAACAA GAGTCCACTA
 5741 TAAAAAGACG TGGACTCCAA CGTCAAAAGG CGAAAACCTC TTATCAGGG CGATGCCCA CTACGTGAAAC
 5811 CATCACCTTA ATCAATTTT TTGGGCTGA GGTGCGCTAA AGCACTAAAT CGGAACCCCTA AAGGGAGCCC
 5881 CCGATTAGA GCTTGACGGG GAAAGCCGGC GAAACTGGGG AGAAGGAGG GGAAGAAAGC GAAAGGAGGG
 5951 GGCCTAGGG CGCTGCCAG TGAGCGCTC ACCTGCGGOG TAAACCAAC ACCGGCCGG CTTAATGCC
 6021 CGCTACAGGG CGCGTCCCG AGTGGCACTT TCGGGAAAT GTGGCGGAA CCCCTATTTG TTATTTTTC
 6091 TAAATCATT CAATATGTA TCCGTCATG AGACATAAC CCTGATAAAAT GTTCAATAA TATTGAAAAAA
 6161 GGAAGGTAT GAGTATCA CATTCTCGT TCGGCTTAT CGGCGATTTT GCGCTCTGT
 6231 TTTCGTCAC CGAAGAACCG TGGTGAAGT AAAAGATGCT GAAGATCAGT TGGGTGCAGC AGTGGGTAC
 6301 ATCGTAATGG ATCTCAACAG CGTAAAGATC CTTGAGAGTT TTGCCCCGA AGAACGTTT CCAATGATGA
 6371 GCACTTTAA AGTCTGCTA TGTTGGCGGG TATTATCCG TATTGACGCC GGGAAAGAGC AACTCGGTG
 6441 CGCGCATCAC TATTCTCAGA ATGACTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGGC
 6511 ATGAGCTAA GAGATTATG CAGTGTGGC ATAACCATGA GTGATAACAC TGCGGCCAAC TTACTCTGA
 6581 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTGTCA CAACATGGGG GATCATGTAA CTGGCTTGA
 6651 TCGTTGGAA CGGGAGCTGA ATGAGCCAT ACCAACCGAC GAGGGTGACA CCACCGATGCC TGAGCAATG
 6721 GCAACACGGT TCGGCAACT ATTAACTGGC GAACACTTCA CTCTAGCTTC CGGCCAACAA TTATAGACT
 6791 GGATGGAGGC GGATAAAGTT GCAGGACAC TTCTGGCTC GGGCTTCCCG GCTGGCTGGT TTATGCTGA
 6861 TAAATCTGGA GCGGGTGAGC GTGGGTCTCG CGGTATCATT CGAGCACTGG GGGCAGATGG TAAAGCCCTCC
 6931 CGTATGCTAG TTATCTACAC GAGGGGGAGT CAGGCAACTA TGGATGAAG AATAGACAG ATGCTGAGA
 7001 TAGGTGCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACTTT AGATTGATT
 7071 AAAACTCTAT TTATTAATTAA AAGGATCTA GTGAGAGTC CTTGTGATA ATCTCATGAC CAAATCCCT
 7141 TAACTGAGT TTCTGTTCCA CTGAGCGCTCA GACCCCGTAG AAAAGATCAA AGGATCTCT TGAGATCCCT
 7211 TTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACT ACCGCTACCA GCGGTGGTTT GTTGCCCGA
 7281 TCAAGAGCTA CCAACTCTTT TTCCGAAGGT AACTGGCTC AGCAGAGCGC AGATACAAA TACTGTCCCT

27/43

7351 CTAGTGTAGG CGTAGTTAGG CCACCACTTC AAGAACTCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
7421 TCCCTGTTACC AGTGGCTGCT GCCAGTGGCG ATAAGTCTG TGCTTACCGGG TTGGACTCAA GACGATAGTT
7491 ACCGGATAAG GCGCAGCGGT CGGGCTGAAC GGGGGTTCG TGACACAGC CCAGCTTGGA GCGAACGACC
7561 TACACCGAAC TGAGATACCT ACAGCGTGAG CTATGAGAAA GCGCCACGCT TCCCGAAGGG AGAAAGGCC
7631 ACAGGTATCC GGTAGACGGC AGGTCGGAA CAGGAGAGCG CACGGAGGGAG CTTCAGGGG GAAACGCCCTG
7701 GTATCTTTAT AGTCCTGTCG GGTTCGGCA CCTCTGACTT GACGGTCGAT TTTCCTGATG CTGGTCAGGG
7771 GGGCGGAGGC TATGGAAAAA CGCCAGCAAC GCGGGCTTTTG TAAGGTTCTT GGCCTTTTG TGCCCTTTTG
7841 CTCACATGTT CTTCCTGCG TTATCCCCCTG ATTCTGTTGGG TAACCGTATT ACCGGCTTTG AGTGAGCTGA
7911 TACCGCTCGC CGCACCGGA CGACCGAGCG CAGCGAGTCA GTGAGCGAGG AAGCGGAAGA GCGCCCAATA
7981 CGCAAAACCGC CTCTCCCCCG GCCTTGGCCG ATTCAATTAT GCACGTGGCA CGACAGTTT CCCGACTGG
8051 AAGCGGGCAG TGACCCCAAC GCAATTATAG TGACTTACGGT CACTCAATTAG GCAACCCAGG CTTTACACTT
8121 TATGCTTCCC GCTCCATATGT TGTTGGAAT TGTTGACGGG TAACAATTTC ACACAGGAAR CAGCTATGAC
8191 CATGATTACG CCAAGCTCGG AATTAACCCCT CACTAAAGGG AACAAAAGCT GCTGAGGGT CCCTAACCTG
8261 CAAGCCCCAC AGTGTGCCCT GAGGCTGCCCT CTTCCTTCTA GCGGCTGCCCT CCACCTGGGT TTGCTTCCC
8331 TAGTTTCACT TACTTGGTT CAGCCAAGGT CTGAAACTAG GTGGCACAG AGCGGTAAGA CTGGGAGAGA
8401 AAGAGACCGAG CTTCACAGGG GGTTTATCAC AGTGCACCCCT GACAGTCCTT AGCCTCACAG GGGGTTTATC
8471 ACATTCACCC CTGACAGTCG TCAGCCCTCAC AGGGGGTTTA TCACAGTGCA CCCTTACAAT CATTCCATTT
8541 GATTCAAAAT TTTTTTAGTC TCTACTGTGCT CTAACCTTGTA AGTTAAATTG GATCAGAGGT GTGTTCCCAG
8611 AGGGGAAAAC AGTATATACA GGGTTCAGTA CTATGCAATT TCAGGCCCTCC ACCCTGGGTCT TGGAAATGTGT
8681 CCCCCGAGGG GTGATGACTA CCTCAGTTGG ATCTCCACAG GTCACAGTGA CACAGATAAA CCAAGACACC
8751 TCCCAAGGCT ACCACAATGG GCGGCCCTCC ACGTGCACAT GCGGGAGGA ACTGCCATGTG CGGAGGTGCA
8821 AGCACACCTG CGCATCAGAG TCTTGGTGT GGAGGGAGGG ACCAGCGCAG CTTCAGGCCA TCCACCTGAT
8891 GAACAGAACCC TAGGGAAAGC CCCAGTTCTA CTTACACCCAG GAAAGGC

28/43

FIGURE 11

1 TGGAAAGGGCT AATTGGTCC CAAAAAAGAC AAGAGATCCT TGATCTGTGG ATCTACCACA CACAAGGCTA
 71 CTTCCTGTAT TGGCAGAACT ACACACCAGG GCCAGGGATC AGATATCCAC TGACCTTTGG ATGGTGCTTC
 141 AAGTTAGTAC CAGTGAACC AGAGCAAGTA GAAGAGGCCA AATAAGGAGA GAAGAACACC TTGTTACACC
 211 CTATGAGCCA GCATGGGATG GAGGACCCGG AGGGAGAACT ATTAGTGTGG AAGTTTGACA GCCTCTAGC
 281 ATTTGTCAC ATGGCCCCAG AGCTGCATCC GGAGTACTAC AAAGACTGCT GACATCGGCC TTTCTACAAG
 351 GGACTTTCGG CTGGGGACTT TCCAGGGAGG TGTTGGCTTG GCGGGACTCG GGAGTGGCGA GCCTCTAGAT
 421 GCTACATATA ACCAGCTGCT TTTGCGCTGT ACTGGGTCTC TCTGGTTAGA CCAGATCTGA GCCTGGGAGC
 491 TCTCTGGCTA ACTAGGGAAC CCACTGCTTA AGCCTCAATA AAGCTTGCTT TGAGTGTCTA AAGTAGTGTG
 561 TGCCCCTCTG TTGTGTACT CTGGTAACTA GAGATCCCTC AGACCCCTTT AGTCAGTGTG GAAAATCTCT
 631 AGCAGTGGCG CCCGAACAGG GACTTGAAG CGAAAGTAAA CCCAGAGGAG ATCTCTGAC GCAGGACTCG

BssHII (711)

701 GCTTGCTGAA CGCGCacgg caagggcggg cggggggcgC cggACggGg ccggcaaaaaat ttgtgactagg

Clal (830)

771 ggacggctaga agggagaaC TCGTGGGG AGCTTCAGTA TAAACGGGG GGGTTTGA TGATGGGAA

841 AAAATTGGGT TAAAGCCAGG GGGAAAGAG AAGTACAAGG TAAAGGAAAT COTATGGCA ACCAGGGAGC

AccI (959)

911 TAGAAGGATT CGCAGTTATC CTGCGCTGT TAGAAACATC AGAAGGCTGT AGACAAATAC TGGGACAGCT

981 ACAACCCATCC CTTAGACAG GATCAGAGA GCTGGATCA CTATACACA CAGTAGGAAAC CCTCTATTGT

1051 GTGCACCCAGC GGATCGAGAT CAAGGACACC AAGGGAGCT TAGACAAGT AGGGGAGAG CAAACAAAGT

1121 CCAAGAGAA GGGCGAGGAG GCGCGAGCTG ACAGGGAGC CAGCAATCAG GTCAGGCAA ATTACCCAT

1191 ATGGCAGAAC ATCCAGGGGC AAATGGTACA TCGGCGATA TCACTAGAA CTTTAACGA TAAAGCTTGGG

1261 AGTTCCGGT TACATAACTT ACGSTAAATG GCCCGCCTGG CTGACGGCCC AACGACCCCCC GCCCATTTGAC

1331 GTCATAATG ACCTATGTTT CCATAGTAAAC GCGCATAGGG ACTTTCCATT GACGTCAATG GGTGGAGTAT

1401 TTACGGTAAA CTGCCCCATTG GCGAGTACAT CAAGTGTATC ATATGCCAAG TACGCCCCCT ATTGACGTCA

1471 ATGACGGTAA ATGGCCCGCC TGGCAATTG CCCAGTACAT GACCTTATGG GACTTTCTA CTTGGCAGTA

1541 CATCTACGTA TTAGTCATCG CTATACCAT GGTTGATGGGG TTTGGCACT ACATCAATGG GCGTGGATAG

29/43

1611 CGGTTTGACT CACGGGGATT TCCAAGTC TCACCCCATG AGTCAATGG GAGTTTGT TTGCAACAAA
 1681 ATCAACGGGA CTTTCCAAA TGTGTAACA ACTTCCGCCCC ATTGAGGCA ATGGGGGGA GCGGTGTACG
 1751 GTGGGAGGTC TATATAAGCA GAGCTCGTT AGTGAACGT CAGATGCCG GGAGAAGGCCA TCCACGGCTGT
 1821 TTTGACCTCC ATAGAAGACA CGAAGCTCTAG AG^{gatacc}ATC TAAGTAACTG TGGCATTCGG CTACTGTTGG
 1891 TAAATGGAA GACGCCAAA ACATAAAGAA AGGCCCGGG CCATCTTATC CTCTAGAGGA TGGAAACCGCT
 1961 GGAGAGCAAC TSCATAAGGC TATGAAGAGA TACGCCCTGG TTCCCTGGAAC AATGGGTTT AAGAGATCAC
 2031 ATATCGAGGT GAACATCACG TACCGGAAT ACTTCGAAT GTCCGTTTCGG TTGGCAGAAG CTATGAAACG
 2101 ATATGGGCTG AATACAAATC ACAGAACTGT COTATGCCGT GAAAACCTCTC TTCAATTCTT TATGCCCGTG
 2171 TTGGGCCGT TATTTATCGG AGTTCAGT GGGCCCGGA AGCACATTAA TAATGAACGT GATTTGCTCA
 2241 ACAGTATGAA CATTTCGAG CCTACCGTAG TTTTGTTC CAARAGGGG TTGCAAAAAA TTTTGAACGT
 2311 GCAAAAAAA TTACCAATAA TCCAGAAAAT TATTATCATG GATTCTAAA CGGATTACCA GGGATTTCA
 2381 TCGATGTACA CGTTCTGCAC ATCTCATCTA CCTCCGGTT TTAAATGAATA CGATTTGTAA CGAGAGTCCT
 2451 TTGATCGTGA CAAACAAATT GCACTGATAA TGATTCCTC TGGATCTACT GGTTTACCTA AGGGTGTGGC
 2521 CCTTCGGCAT AGAACCTGCT GCGTCAAGTT CTGGCATGCC AGAGATCTA TTTTTGGCAA TCAAATCATT
 2591 CCGGATACTG CGATTTAAAG TGTGTTGCA TTGATCGACG GTTTTGGAAAT GTTTTACTACA CTCGGATATT
 2661 TGATATGTGG ATTTCGAGTC GTCTTAAGT ATAGTTTGA AGAAGAGCTG TTTTTAGGT CCTTTCAGGA
 2731 TTACAAATT CAAAGTGCCT TGCTAGTACG AACCTTATTT TCTTCTTCG CCAAAAGCAC TCTGATTGAC
 2801 AAATACGATT TATCTTAATT ACACGAAATT GCTTCTGGGG GCGCAGCTCT TTCGAAAGAA GTGGGGGAAG
 2871 CGGTTCCAAA ACGCTTCAT CTGGCTGGGA TACGCTCTGG ATATGGGCTC ACTGAGACTA CATCAGCTAT
 2941 TCTGATTACG CCCGAGGGGG ATGATAACC GGGGGGGGTG GGTAAAGTIG TTCCATTGTTT TGAAGGGAAAG
 3011 GTTGTGGATC TGGATACCGG GAAAACGCTG GGCGTTAAATC AGAGAGGGGA ATTATGTGTC AGAGGACCTA
 3081 TGATTTATGTC CGGTTATGTA AACATACGGG AAGGAGCCAA CGGCTTGATT GACAAGGTG GATGGCTACA
 3151 TTCTGGAGAC ATAGCTTAATC GGGACGAGA CGAACACTTC TICATAGTTG ACCGCTTGAA GTCTTTAATT

ClaI (3259)

3221 AAATACAAAG GATATCAGGT GGCCCCCGGT GAAATTGAAAT CGATATTGTT ACAACACCCC AACATCTTCG
 3291 ACGCGGGCGT GGCAGGTCTT CGCGAGGATG AGCGCCGCTGA ACTTCCCGCC GCCGTTGTIG TTTTGGAGCA
 3361 CGGAAAGACG ATGAGCGGAAAT AAGGAGCTG GGATTACGTC CGCAAGTCAG TAACACCCG GAAAAGCTG
 3431 CGCGGAGGAG TTGTGTTGT GGACGAAGTA CGGAAAGCTC TTACCGGAAA ACTCGACGCCA AGAAAAAATCA

30/43

Apal (3557)

Xhol (3548) KpnI (3563)

3501 GAGAGATCCT CATAAGGCC AAGAAGGGCG GAAAGTCCAA ATTCATACTC GAGGGGGGGC CGGGTACCTT
 3571 TAAGACCAAT GACTTACAAAG GCAGCTGTAG ATCTTAGCCA CTTTTTAAA GAAAAGGGGG GACTGGAAGG
 3641 GCTAAATTCAAC TCCCAGAGAA GACAAGATAT CCTTGATCTG TGATCTACC ACACACAGG CTACTTCCCT
 3711 GATTGGCAGA ACTACACACC AGGGCCAGGG GTCAAGATATC CACTGACCTT TGGATGGTGC TACAAGCTAG
 3781 TACCAAGTGA GCCAGATAAG GTAGAAGAGG CCAATAAAGG AGAGAACACCC AGCTTGTAC ACCCTGTGAG
 3851 CCTGCATGGA ATGGATGACC CTGAGAGAGA AGTGTAGAG TGGAGGTTTG ACAGCCGCT ACCATTTCAT
 3921 CACGTGGCCC GAGAGCTGCA TCCGGAGTAC TTCAAGAACT GCTGACATCG AGCTTGTAC AAGGGACTTT
 3991 CCGCTGGGGA CTTTCCAGGG AGGGGTGGCC TGGGGGGAC TGGGGAGTGG CGAGCCCTCA GATGCTGCAT
 4061 ATAAGCAGCT GCTTTTGCC TGTAATGGGT CTCTCTGGTT AGACCAAGATC TGAGCCCTGGG AGCTCTCTGG
 4131 CTAACCTAGGG AACCCACTGC TTAAAGCTCA ATAAGCTTG CCTGAGTGC TTCAAGTAGT GTGTGCCCGT
 4201 CTGTTGTGTG ACTCTGGTAA CTAGAGATCC CTCAGACCTT TTAGTCAGT GTGGAAATC TCTAGCACCC
 4271 CCCAGGAGGT AGAGGTGCA GTGAGGCCAG ATCGGCCAC TGCATTCCAG CCTGGCAAG AAAACAAGAC
 4341 TGTCTAAAT ATAATAATA AGTAAAGGT ATTAAATATA TTATACATG GAGGTATAA AAATATATAT
 4411 ATTTGGCTG GGCGCAGTGG CTACACCTG CGCCCGGGCC TTGGGGAGGC CGAGGCAGGT GGATCACCTG
 4481 AGTTGGAG TTCCAGACCA GCTTGACCAA CATGGAGAAA CCCTCTCTCT GTGTATTTTT AGTAGATTIT
 4551 ATTTATGTG TATTTTATTC ACAGGTATT CTGGAAAATC GAAACTGTTT TTCTCTACT CTGATACAC
 4621 AAGAATCATC AGCACAGAGG AAGACTTCTG TGATCAATG TGTTGGAGA GGGAGGTTT CACCAAGCACA
 4691 TGAGCAGTCA GTTCTGCCCG AGACTCGGGG GTGTCCTTTC GTGTCAGTTC CACACCGGCC TGGCTGGAGA
 4761 GAGGTGAGAC CACAGGGTGA GGGCTCAGTC CCCAAGACAT AAACACCCAA GACATAACCA CCCAACAGGT
 4831 CCACCCGGCC TGCTCCCCAG GCAGAGCGGA TTCAACAGA CGGGAAATTAG GATAGAGAAA GAGTAAGTCA
 4901 CACAGGCC GCTGTGGGG AGAACGGAGT TCTATTATGA CTCAAATCG TCTCCCCAG CATTGGGGGA
 4971 TCAGAGTTT TAAGGATAAC TTAGTGTGA GGGGGCCAGT GAGTTGGAGA TGAAGGGTA GGGAGTCGAA
 5041 GGTGTCCTT TGCGGGAGT CAGTCTCTG GTGGGGGCCA CGACATCGGA TGAGCCAGTT TATCAATCCG
 5111 GGGGTGCGAG CTGATCCATG GAGTGCAGGG TGTGAAAT ATCTCAAGCA CTGATTGATC TTAGGTTTTA
 5181 CAATAGTGAT GTTACCCAG GAACTTTTG GGGAAAGTC GAACTTGTG GCTCTAGTCG GCATGACTCC
 5251 TAAACCATAA TTCTCTTTTT TTGTTTTTTT TTGTTTTTTT GAGACAGGGT CTCACTCTGT CACCTAGGCT
 5321 GGAGTCAGT GGTGCAATCA CAGCTCAGT CAGCCCTAG AGCCGGCTAG ACCGGGGTGG AGCTCCAATT
 5391 CGCCCTTATAG TGACTGTAT TACATTCAC TGCCCCCTGT TTACAACGT CGTGACTTGGG AAAACCTGG
 5461 CGTACCCAA CTTAAATCGCC TTGAGCACCA TCCCCCTTTC GCAAGCTGGC GTAAATAGCGA AGAGGCCCGC
 5531 ACCGATCGCC CTTCCCAACG GTTGGCGAGC CTGAATGGCG AATGGCGGA ATTGTAAAC GTAAATATIT
 5601 TGTTAAATT CGCGTTAAAT TTGTTTTAA TGAGTCATT TTGTTAACCAA TAGGGGGAAA TGGGAAATAAAT
 5671 CCCTTATAAA TCAAAAGAAT AGACCGAGAT AGGTTGAGT GTGTTGGAG CTTGGAACAA GAGTCCACTA
 5741 TAAAGAACG TGACTCTCAA CGTCAAAGGG CGAAACCCCG TCTATCAGG CGATGGCCCA CTACGTGAAC
 5811 CATCACCCCA ATCAAGTTT TTGGGGTOGA GTGCGCGTA AGCACTAAAT CGAACCCCTA AGGGGAGCCC
 5881 CCGATTAGA GCTTGACGGG GAAAGCCCGC GAACGTGGCG AGAAAGGAAG GGAAGAAAGC GAAAGGAGCG
 5951 GGGCTAGGG CGCTGCAGAAG TGAGTCGGTC AGCTCTCCGC TAACCAACCC ACCGGCGGCC CTTAAATGCC
 6021 CGCTACAGGG CGCGCTCCAG GTGGCACTT TCGGGAAAT GTGCGGGAAAC CCCTTATTG TTATTTTTC
 6091 TAAATACATT CAAATATGTA TCCCTCTATG AGACATAAC CCTGATAAAT GCTTCATAAA TATTGAAAAA
 6161 GGAAGACTAT GAGTATCCTA CATTCCGTG TCGCCCTTAT TCCCTTTTTT GGGCATTTT GCGCTCTGT
 6231 TTGTCAC CCAGAACGC TGGTGAAGT AAAAGATGCT GAAGATCAGT TGGGTGCAAG AGTGGGGTAC
 6301 ATCGAATGCA ATCTCAACAG CGGTAAAGTC CTGAGAGTT TTGCCCCCGA AGAACGTTT CCAATGATGA
 6371 GCACTTTAA AGTCTGCTA TGTTGGGGGG TATTTCTCCG TATTTGACCC GGGCAAGAGC AACTGGGTG
 6441 CGCGATACAC TATTCAGA ATGACTTGGT TGAGTACTCA CCAGTCACAG AAAAGCATCT TACGGATGCC
 6511 ATGACAGTAA GAGAATTATG CAGTGCCTGCC ATAACCATGA TGATAAACAC TCGGGCAAC TTACTTCTGA
 6581 CAACGATCGG AGGACCGAAG GAGCTAACCG CTTTTTGCA CAACATGGGG GATCATGTAA CTGGCTTGA
 6651 TCGTGGGAA CGGGAGCTGA ATGAAGCCAT ACCAACGAC GAGCGTGACA CCAGGATGCC TGTAGCAATG

31/43

6721 GCAACRACGT TGGCAAACG ATTAACGGC GAACTACTTA CTCTAGCTTC CGGGCAACAA TTAATAGACT
 6791 GGATGGAGGC GGATAAAGTT GCAGGACAC TCTCTGGCTC GGCCTTCGG CCTGGCTGGT TTATTGCTGA
 6861 TAAATCTGGA GCCGGTGAGC GTGGGTCTCG CGGTATCATC GCACCACTGG GGCAGATGG TAAGCCCTCC
 6931 CGTATCTAG TTATCTACAC GACGGGGAGT CAGGCAACTA TGGATGAACG AAATAGACAG ATCGCTGAGA
 7001 TAGGTGCTC ACTGATTAAG CATTGGTAAC TGTCAGACCA AGTTTACTCA TATATACCTT AGATIGATTT
 7071 AAAACTTCAT TTTTAAATTAA AAGGATCTA GGTGAAGATC CTTTTGATA ATCTCATGAC CAAATCCCT
 7141 TAACGTGAGT TTCTGTTCCA CTGAGCGCTA GACCCCGTAG AAAAGATCAA AGGATCTTCT TGAGATCCCT
 7211 TTTTCTGCG CGTAATCTGC TGCTTGCAAA CAAAAAAACC ACCTACCA GGGTGGTTT GTTIGCCGG
 7281 TCAAGAGCTA CCAACTTCTT TTCCGAAGGT AACTGGCTTC AGCAGAGGG AGATACAAA TACTGTCTT
 7351 CTAGTGTAGC CGTAGTTAGG CCACCACTTC AAGAACCTG TAGCACCGCC TACATACCTC GCTCTGCTAA
 7421 TCTCTTACCG AGTGGCTGCT GCGAGTGGCG ATAAGTCGTG TCTTACCGGG TTGGACTCAA GACGATAGTT
 7491 ACCGGATAAG CGCGAGCGGT CGGGCTGAAC GGGGGTTTCC TGACACAGC CCAGCTTGGG GCGAACGACC
 7561 TACACCGAAC TGAGATACCTT ACAGCGTGAG CTATGAGAAA GCGGACCGCT TCCCGAAGGG AGAAAGGCC
 7631 ACAGGTATCC GTTAAGCGGC AGGGTCGGAA CAGGAGAGGG CAGGAGGGAG 7701 GTATCTTTAT AGTCTCTGCG GTTTCGCA CCTCTGACTT GAGCCTGCGAT TTTTGTGATG CTCGTCAAGGG
 7771 GGGCGGAGCC TATGGAAAAC CGCAGCAAC TACGGTTTCTC GGGCTTTTGC TGGCTTTTGC
 7841 CTCACATGTT CTTCCTGCG TTATCCCTG ATTCCTGCGA TAACCGTATT ACCGGCTTTC AGTGAGCTGA
 7911 TACCCCTCGC CGCAGCGGAA CGACCGAGCC CAGCGAGTCG GTGAGGGAGG AGCGGAAGA GCGCCCAATA
 7981 CGCAAACCGC CTCTCCCCGC CGCTTGGCCG ATTCAATTAT GCAGCTGGCA CGACAGGTTT CCCCACCTGA
 8051 AAGCCGGCAG TGAGCGCAAC GCAATTAAATG TGAGTTAGCT CACTCTATTAG GCACCCCAAGG CTTTACACTT
 8121 TATGTTCCG GCTCGTATGT TGTTGGAAAT TGAGGGCGGA TACAATTTC ACACAGGAAA CAGCTATGAC
 8191 CATGATTAGG CCAAGCTGGG AATTAACCCCT CACTAAAGGG AACAAAGCT CTGCAAGGT CCCTAACCTGC
 8261 CAAGCCCCAC AGTGTGCCT GAGGGCTGCCCTTCTCTCTA GGGGCTGCC CCACCTGGCT TTGCTTTCCC
 8331 TAGTTTCAGT TACTTGGCTT CAGCCCTAGGT CTGAAACTAG GTGGCAGACAG AGCGGTAAGA CTGGAGAGA
 8401 AAGAGACCG AGTGTACAGGG GTTTCATCAG AGTGCACCCCT GACAGTGTG AGCCTCACAG GGGTTTATC
 8471 ACATTCACCC CTGACAGTCG TCAGCTCAGC AGGGGGTTTA TCACAGTGCA CCCTTACAAAT CATTCCATT
 8541 GATTACACAT TTTTTAGTC TCTACTGTGCT CTAACCTGTA AGTTAAATTG GATCAGAGGT GTGTTCCAG
 8611 AGGGAAAAC AGTATATAACA GGGTTCACTA CTATGCATT TGAGGCTCC ACCTGGGTCT TGGAATGTGT
 8681 CCCCCGGAGGG GTGATGACTA CCTCAGTTGG ATCTCCACAG GTCACTGTA CACAAGATAA CCAAGACACC
 8751 TCCCAAGGCT ACCACAAATGG GCGGGCTCC ACCTGCACAT GCGGGAGGG ACTGCCATGT CGGAGGTGCA
 8821 AGCACACCTG CGCATCAGAG TCTCTGGTGT GGAGGGAGGG ACCAGCGCAG CTTCCAGGCA TCCACCTGAT
 8891 GAACAGAACC TAGGGAAGC CCCAGTCTA CTACACCCAG GAAAGGC

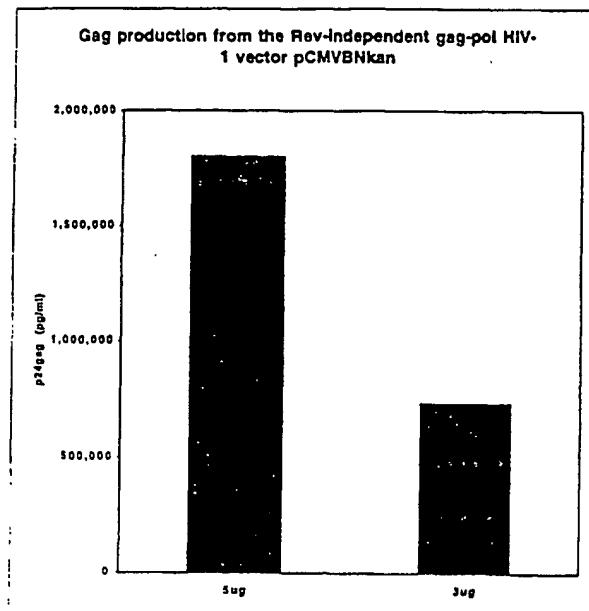
32/43

FIGURE 12

mBCwCN frag	-----	-----	C-	-AC-	-G-	-----
m2BCwCN frag	-----	-----	C-	-G-	-G-	-----
BC/HXB2	-----	-----				-----
BC/NL43	-----	-----				-----
#1				
	CGCGCACGGC	AAGAGGCAG	GGGCAGGCAC	TGGTGAGTAC	GCCAAAAATT	
mBCwCN frag	-----	-----	C-	C-	-----	-----
m2BCwCN frag	-----	-----				-----
BC/HXB2	-----	T-	-----	-----	-----	-----
BC/NL43	-----	-----	-----	-----	G-	-----
#51	
	TTGACTAGCG	GAGGCTAGAA	GGAGAGAGAT	<u>GGGTGCGAGA</u>	GCGTCAGTAT	
mBCwCN frag	-----	-----				-----
m2BCwCN frag	-----	-----				-----
BC/HXB2	-----	-----				-----
BC/NL43	-----	-----	AA-	-----		-----
#101		
	TAAGCGGGGG	AGAATTAGAT	CG			

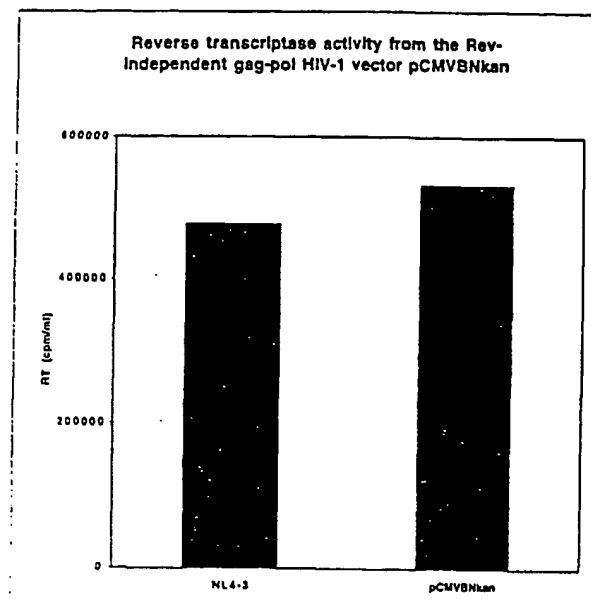
33/43

FIGURE 13

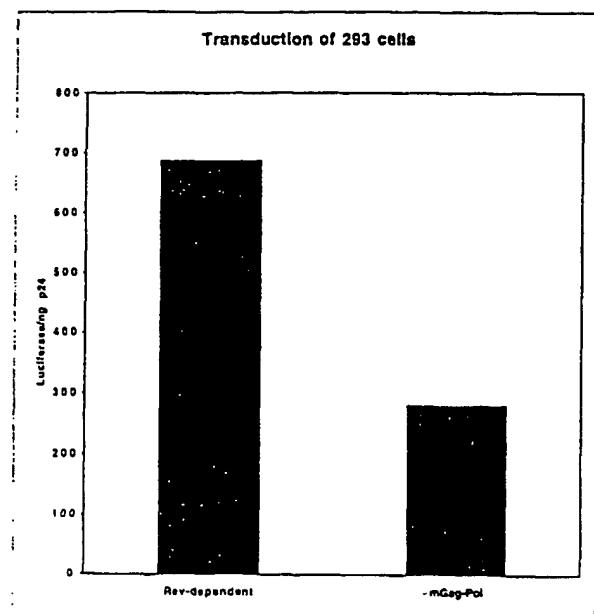


34/43

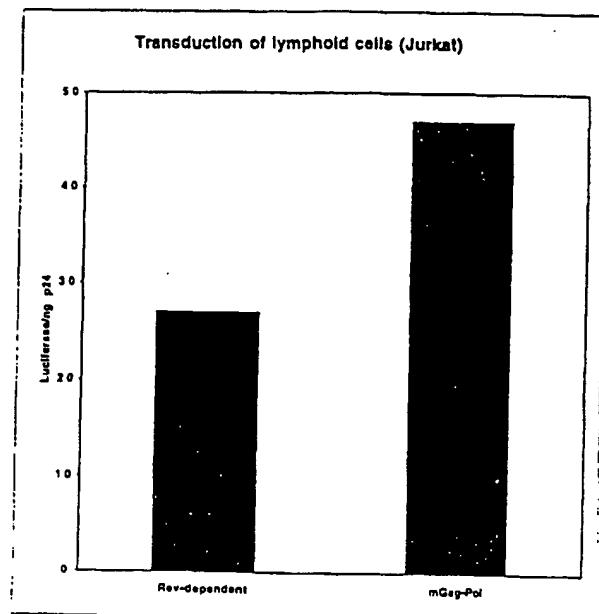
FIGURE 14



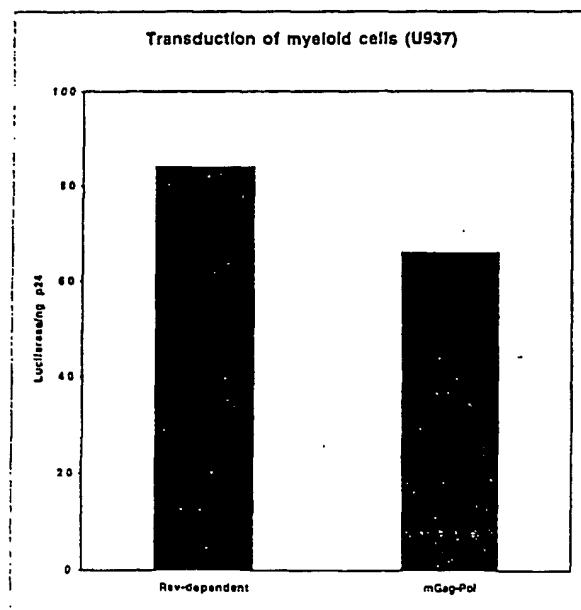
35/43

FIGURE 15A

36/43

FIGURE 15B

37/43

FIGURE 15C

38/43

FIGURE 15D

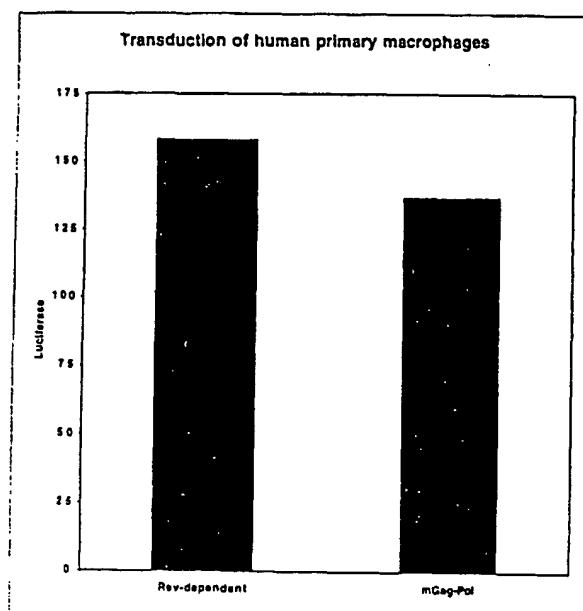


FIGURE 16

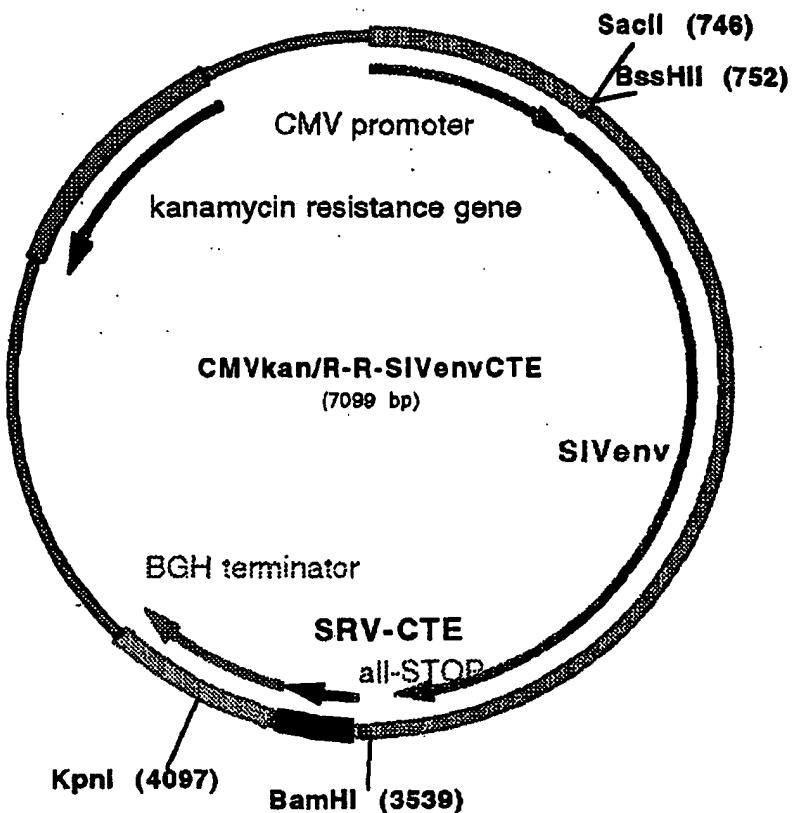


FIGURE 17

1 CCTGGCCATT GCATACTTGT TATCCATATC ATAATATGTA CATTATATTGTC GCTCATGTC CAACATTAC
 71 GCCATGTGA CATTGATTAT TGACTAGTTA TTAATAGTA TCAATTACGG GGTCAATTAGT TCATAGCC
 141 TATATGGAGT TCCGGCTTAC ATAACCTTACG GTAAATGGCC CGCCCTGGTG ACGGCCAAC GACCCCCGCG
 211 CATTGACGTC AATAATGACG TAATGTCACA TAGTAACGCG AATAGGGACT TTGCAATTGAC GTCAATGGGT
 281 GGAGTATTTA CGGTTAACTG CCCACCTGGC AGTACATCAA GTGTAATCATA TCCCAAGTAC GCCCCCTATT
 351 GACCTCAATG ACGGTAAATG GCGCGCTTGG CATTATGCCG AGTACATGAC CTTATGGAC TTTCTACTT
 421 GGCAGTACAT CTACGTTATTG TGCATCGCTA TTACCATGGT GATGGGGTIT TGGCAATGACA TCAATGGGG
 491 TGGATAGCGG TTTGACTCAC GGGGATTTCG AAGTCTCCAC CCCATTGACG TCAATGGAG TTTGTTTTCG
 561 CACCAAAATC AACGGGACTT TCCAAAATGT CGTAACAACT CGGCGCTTATT GACCCAAATG GCGCGTAGGC
 631 GTGTAACGGTG CGAGGCTAT ATAAGCAGAG CTGGTTTAST GAAACGTCAG ATCGCGTGGG GACCCATCC

BssHII (752)**SacII (746)**

701 ACGCTGTTT GACCTCATA GAAGACACCG GCGACGATCC AGCCTCCCGG GCGCGCTAA GTATGGGATG 1► M G C
 771 TCTTGGGAAT CAGCTGCTTA TGGCCATCTT GCTTTTAAGT GCTCTAAGGAA TCTTATTGTCAC TCTTATAATGTC
 3► L G N Q L L I A I L L S V Y G I Y C T L Y V
 841 ACACTCTTTT ATGCTGTACC AGCTTGGAGG AATGGACAA TCCCTCTT TCTGCGAACCC RAGAATAGGG
 27► T V F Y G V P A W R N A T I P L F C A T K N R
 911 ATACITGGGG AACAATCAG TGCCTTACCG ATAATGGTG AATTCAGAA GTGGCCCTTA ATGTTACAGA
 56► D T W G T T Q C L P D N G D Y S E V A L N V T E
 981 AACCTTGTGCTTGGAAATA ATGACTCAC AGAACACCC AATAGGAGT TATGGCGACT CTTCAGACX
 73► S F D A W N N T V T E Q A I E D V W Q L F E T
 1051 TCAATAAAGC CTGGTGTAAA ATTATGCCA TTATGCTTAA CTATGAGATG CAATAAAAGT GAGACAGATA
 97► S I K P C V K L S P L C I T M R C N K S E T D
 1121 GATGGGATT GACRAAATCA ATARACACRA CAGCATCAC AACATCACG ACACATCACG CAAACTAGA
 120► R W G L T K S I T T T A S T T S T T A S A K V D
 1191 CATGGGAAAT GAGACTAGTT CTGGTATAGC CCAGGATANT TGCACAGGCT TGGAAACAAGA GCAATGATA
 143► M V N E T S S C I A Q D N C T G L E Q E Q M I
 1261 AGCTGTAAAT TCAACATCAC AGGTTAAAAA AGAGACAAGA AAAAGAGTA CAATGAAACT TGGTACCTG
 167► S C K F N M T G L K R O K K K E Y N E T W Y S
 1331 CACATTGTGAT GTGAAACAA CGCAATAACA CTGGTAATGA AAGTAGATGT PACATGAACC ACTGTAAACAC
 190► A D L V C E Q G N N T G N E S R C Y M N H C N I
 1401 TTCTGTATC CAAGACTTGT GTGACAAACA TTATGGGAT GCTATTAGAT TTAGGTATG TGCACCTGCA
 213► S V I Q E S C D K H Y W D A I R F R Y C A P P
 1471 GGTTATGCTT TGCTTATGAT TAATGACACA AATTATTCAG GCTTATGCC TAAATGTCI AAGGTTGGG
 237► G Y A L L R C N D T N Y S G F M P K G S K V V
 1541 TCTCTTCATG CACAGGATG ATGGAGACAC AGACCTCTAC TTGGTTGGC TTATGGAA CTAGACCGA
 260► V S S C T R M M E T Q T S T W F G F N G T R A E
 1611 AAATAGAACT TATATTTACT GGCATGGTAC CGATAATAGC ACTATAATTA GTTAAATAA GTATTATAAT
 283► N R T Y I Y W H G H D N R T I I S L N K Y Y N
 1681 CTAACATGA AATGTAGAAG ACCAGGAAT AAGACAGTT TACCACTCAC CATTATGTCI GGATGGTTT
 307► L T M K C R R P G N K T V L P V T I M S G L V
 1751 TCCACTCAC ACCATCAAT GTAGGCCA AGCAGGCCATG GTGTTGGTTT GGAGGAAAT GGAAGGAAAC
 330► F H S Q P I N D R P K Q A W C W F G G K W K D A
 1821 AATAAAAGAG GTGAACAGA CCTTGTCAA ACATUCCAGG TATAGTGAA CTAACAAATC TGATANATC
 353► I K E V K O T I V K H P R Y T G T N N T D K I
 1891 AATTTGACCG CTCCTGGAGG AGGAGATCCG GAAGTTACCT TCATGNGAC AAAATGCCAGA CGAGAGTC
 377► N L T A P G G G D P E V T F M W T N C R G E F

1961 TCTACTGTAA AATGAAATGG TTTCTTAAATG CGGTAGAACAGA TAGGGATACA CCTTAACCGAG ACCCAAAGGA
 400>L Y C K M N W F L N W V E D R N T A N Q K P K E
 2031 ACAGCATAAA AGGAAATTAG TGCCATGTCA TATTAAGACAA ATAATCAACA CTTCGCATAA ACTTGGCAAA
 423>Q H K R N Y V P C H I R Q I I N T W H K V G K
 2101 AATGTTTATT TGCCCTCAAC AGAAGGAGAC CTCACGTGTA ACTTCACAGT GACCAAGTCTC ATAGCAAACA
 447>N V Y L P P R E G D L T C N S T V T S L I A N
 2171 TAGATGGAT TGATGGAAAC CAAACTAATA TCACCATGAG TGCAAGGGTG GCAGAACTGT ATGGATTGGA
 470>I D W I D G N Q T N I T M S A E V A E L Y R L E
 2241 ATTGGGAGAT TATAAATTAG TAGAGATCAC TCCAAATTGCG TTGGGCCCA CAGATGTGAA GAGGTACACT
 493>L G D Y K L V E I T P I G L A P T D V K R Y T
 2311 ACTGGTGGCA CCTCAAGAAA TAAAGAGGGG GCTTGTGTC TACGGMTCTT GGGTTTTCCTC GCAACGGCAG
 517>T G G T S R N K R G V F V L G F L G F L A T A
 2381 GTTCTGCAAT GGGAGGCC AGCCTGACCC TCAOGGCACA GTCCGAACT TTATTGCTG GGATAGTCCA
 540>G S A M G A A S L T L T A Q S R T L L A G I V Q
 2451 ACAGCAGCAA CAGCTGTGG AGCTGGTCAA GAGACAACAA GAACTTGTG GACTGACCTG CTGGGAAACA
 563>Q Q Q Q L L D V V K R Q O E L L R L T V W G T
 2521 AAGAACCTCC AGACTTGGT CACTGCCATC GAGAACTACT TAAAGGACCA GGCGCAGCTG ATGCTTGGG
 587>K N L Q T R V T A I E K Y L K D Q A Q L N A W
 2591 GATGTGGTT TAGACAAGTC TGCCACACTA CTGTACCATG GCCAAATCCA AGCTAAACAC CAAAGTGGAA
 610>G C A F R Q V C H T T V P W P N A S L T P K W N
 2661 CAATGAGACT TGGCAAGATG CGGAGCGAAA GGTTGACITC TTGGAAGAAA ATATAACAGC CCCTCTAGAC
 633>N E T W Q E W E R K V D F L E E N I T A L L E
 2731 GAGGCACAAA TTCACAAAGA GAAGAACATG TAGAAATTG AAAAGTGTAA TAGCTGGAT GTGTTGGCA
 657>E A Q I Q Q E K N M Y E L Q K L N S W D V F G
 2801 ATGGTTTGA CCTTGTCTCT TGGATAAAAGT ATATAACAATA TGGAGTTTAT ATAGTTGTAG GAGTAATACT
 680>N W F D L A S W I K Y I Q Y G V Y I V V G V I L
 2871 GTTAAGAATA GTGATCTATA TAGTACAAAT CCTAGCTAAG TTAAGGCAGG GGTATAGGCC AGTGTCTCT
 703>L R I V I Y I V Q M L A K L R Q G Y R P V F S
 2941 TCCCCACCTT CCTTATTCCA GCAGACCCAT ATCCAACAGS ACCCGGCACT GCAACCCAGA GAAGGCAAAAC
 727>S P P S Y F Q Q T H I Q Q D P A L P T R E G K
 3011 AAAGAGACGG TGGAGAAGGC GGTGCAACA CCTCTGGGC TTGGCAGATA GAAATAATCC ACTTTCCTAT
 750>E R D G G E G G G N S S W P W Q I E Y I H F L I
 3081 TCGTCAGCTT ATTCAGACTCT TGACTTGGCT ATTCAGCTAAC TGTAGGACTT TGTATCGAG ACTATACCG
 773>R Q L I R L L T W L F S N C R T L L S R V Y Q
 3151 ATCCCTCAAC CAATACTCCA GAGGCTCTCT GCGACCCCTAC AGAGGATCC AGAAGTCTC AGGACTGAAAC
 797>I L Q P I L Q R L S A T L Q R I R E V L R T E
 3221 TGACCTACCT ACAATATCGG TGGAGCTATT CCTATGGGC GGTCCAGGCC GTCTGGAGAT CTGGCGACCG
 820>L T Y L Q Y G W S Y F H E A V Q A V W R S G D R
 3291 ACAGAGACTC TGGGGGGCGC GTGGGGGGAC TTATGGGAGA CTCTTAGAGA CTCTTGCGGG CGCGTGGGG
 843>Q R L L R A R G E T Y G R L L E T L A G A W -G
 3361 GCTTATCGG AGACTCTTAG GAGAGGAGAG GTGGAGATG GATACTGGAG GTGGAGATG CATACTGCC
 867>D L W E T L R R G E V E D G Y S R W K M D T R
 3431 AAATCCATCC CCTGGAGGT TAGACAAGGG CCTGAGCTA CTCTCTGTG ACCAGGAGGA TTAGACANGG
 890>Q I H P Q E D

BamHI (3539)

3501 CCTTGAGCTC ACTCTCTCTGT GAGGGACAGA GGGACAGAGgg atccacttagt ttatgaaCTCG AGGGGGGGCC
 3571 CGGTACGAGC CCTTAACTTG CTAGAGACCA CCTCCCTGC GAGCTAACGT GGACAGCCAA TGACGGGTAA
 3641 GAGAGTGACA TTTTCACTA ACCTAACAGA GGAGGGCGT CAGAGCTACT GCTTAATCCA AAGACGGGTA
 3711 AAATGATAA AAATGTATCA CCTCAACCTA AGACAGGCGC AGCTTCCGAG GGATTTGTG TCTGTTTAT
 3781 ATATATTTAA AAGGGGAGACC TGTCGGAGC CGTGCTGCC GGATGATGTC TTGGTCTAGA CTGGAGGGGG
 3851 GGGCGGTAC GATCACATC TGCTGTGCT TCTAGTGTG AGCCATCTGT TTGGTGTGCC TCCCGGTGCG
 3921 CCTCCCTGAC CCTGGAGGT CCCACTCCCA CTCTCTTTC CTATAAATG GAGGAATIG CATGGCATIG

3991 TCTTAACTAGT TGTGATTCCTA TTCTGGGGU TGGGTUUGG CAGGCACACCA AGGGGAAAGGA TTGGGAACAC

KpnI (4097)

4061 AATAGCAGGC ATGCTGGGG A TCGGGTGGGC TCTAATGGTA CGCAGGTGCTP GAAGAAATTGA CGCGGTTCTP
 4131 CCTGAGGCCAU AAAGGAAACCA UCAATTCGU TTTTNTGTA CACACUCCGTP CCAGGTTCTP TTTCCTTAGT
 4201 TCCAGGCCCA CTCTAAGGAA ACTCTAAGCT CAGGAGGCTC CGGGTTCA A TCCAGGCCCA TAAAGTACCT
 4271 CGAGGCTCT CTGGCTCTC CATCAGGCCA CGAACCCAA CGTAGCTCAG AAGAGTGGGA AGAAATTAA
 4341 CGAAGATGG CTATAAAGTG CAGAGGGAGA GAAATCTCCT CGAACATGTC AGGAAGTAAT GACAATAT
 4411 ATAGAAATTTC TTCCGCTTCC TCGCTCACTG ACTCGCTGOG CTCGGTGTGTT CGGCTGGGC GAGCGGTATC
 4481 AGCTCACTCA AAGGGGAA TACGGTTATC CACAGAATCA GGGGATAACG CAGGAAAGAA CATGIGAGCA
 4551 AAAGGCCAGC AAAAGGCCAG GAAACGTAAT AAAGCGCGT TGCTGGCGTT TTTCATAGG CTCCGCCCCC
 4621 CTGACGAGCA TCACAAAAAT CGACGCTCAA GTCACTGGT GCGAAACCCCG ACAGGACTAT AAAGATACCA
 4691 GGGTTTCCC CCTCGGAAGCT CCCTCGTGGG CTCTCTCTGTT CGGACCCCTGC CGCTTACCGG ATACCTGTCC
 4761 GCTTTCCTCC CTTCGGGAAG CGTGGGCGTT TCTCAATGCT CACGCTGTAG GTATCTCAGT TCGGTGTAGG
 4831 TCGTTGCTC CAAGCTGGG TGTTGTGCAAG AAAAAACCGT TCAGCCGAC CGCTGGCCT TATCCGGTAA
 4901 CTATCGTCTT GAGTCACACC CGGTAAAGACA CGACITATCG CCACCTGGCAG CAGCAGTGG TACAGGATT
 4971 ACCAGAGCGA GGTTATGTAGG CGGTGCTACA GAGTTCTTGA AGTGGTGGCC TAACTACGGC TACACTAGAA
 5041 GGACAGTATT TGGTATCTGC GCTCTGCTGA AGCCAGTTAC CTCGGAAAA AGAGTTGGTA CCTCTTGATC
 5111 CGGCAAACAA ACCACCGCTG GTAGCGGTGG TTTTTTTGTT TGCAAGCAGC AGATTACCGC CAGAAAAAAA
 5181 GGATCTCAAG AAGATCTTT GATCTTTCT ACAGGGCTCTG ACGCTCAGT GAACGAAAAC TCACGTTAAG
 5251 GGATTTGGT CATGAGATTA TCAAAAAGGA TCTTCACCTA GATCCTTTA ATTAAAAAT GAAGTTTTAA
 5321 ATCAATCTAA AGTATATATG AGTAAACITG GTCTGACAGT TACCAATGCT TAATCAGTGA GGCACCTATC
 5391 TCAGCGATCT GTCTATTCG TTCACTCCATA GTGGCTGAC TCCGGGGGG GGGGGCGCTG AGGTCTGCCT
 5461 CGTGAAGAG GTGTTGCTGA CTCATACCG CGCTGAATCG CCCCATCATC CAGCCAGAAA GTGAGGGAGC
 5531 CACGGTTGAT GAGAGCTTG TTGTAGGTGG ACCAGTTGGT GATTITGAAC TTTTGTGTTG CCACGGAAAGC
 5601 GTCTCGTGTG TCGGGAAAGAT GCGTGATCTG ATCCCTCAAC TCAGCAAAG TICGATTITAT TCAACAAAGC
 5671 CGCGTCCCCG TCAAGTCAGC GTAATGCTCT CGCAGTGTAA CAACCAATTA ACCAATTCTG ATTAGAAAAAA
 271 F F
 5741 CTCATCGAGC ATCAAATGAA ACTGCAATTG ATTCATATCA GGATTATCAA TACCATATT TTTGAAAAAGC
 269 E D L M L H F Q L K N M D P N D I G Y K Q F L
 5811 CGTTTCTGTA ATGAAGGAGA AAACTCACCG AGGCAGTTTCG ATAGGATGGC AAGATCCCTGG TATCGGTCTG
 245 R K Q L S P S F E G L C N W L I A L D Q Y R D A
 5881 CGATTCCGAC TCGTCCAACA TCAATACAAC CTATTAATTG CCCCCTGGTCA AAAATAAGGT TATCAAGTGA
 222 I G V R G V D I C G I L K G E D F I L N D L S
 5951 GAAATCACCA TGAGTGCAGCA CTGAATCCGG TGAGAATGCC AAAACCTTAT GCATTTCTTT CGAGACTTGT
 199 F D G H T V V S D P S P F L L K H M E K W V Q
 6021 TCAACAGGGCC AGCCATTACG CTGGTCATCA AAATCACTCG CATCAACCAA ACGTTATTC ATTCTGTGATT
 175 E V P W G N R E D D F D S A D V L G N N M R S Q
 6091 GCGCTGAGC GAGACGAAAT ACGCGATCGC TGTTAAAAGG ACAATTACAA ACAGGAATCG AATGCAACCG
 152 A Q A L R F V R D S N F P C N C V P I S H L R
 6161 CGCGAGGAAC ACTGCCAGCG CATCAACAT ATTTCACCT GAATCAGGAT ATTCTCTAA TACCTGGAAAT
 129 R L F V A L A D V I N E G S D P Y E E L V Q F
 6231 GCTGTTTCC CGGGGATCGC AGTGGTGAGT AACCATGCAT CATCAGGAGT ACGGATAAAA TGCTTGATGG
 105 A T K G P I A T T L L W A D D P T R I F H K I T
 6301 TCGGAAGAGG CATAAAATCC GTCACTGGTGTAGCTGCATCATCT GTAAACATCAT TGGCAACGCT
 82 P L P M F E T L W N L R V M E D T V D N A V S
 6371 ACCTTTCGCA TGTTCAGAA ACAACTCTGG CGCATCGGGC TTCCCATACA ATCGATAGAT TGTCGCACT
 59 G K G H K L F L E P A D P K G Y L R Y I T A G
 6441 GATTGCCGA CATTATCGCG AGCCCATTTA TACCCATATA AATCAGCATC CATGTTGGAA TTTAATCGC
 35 S Q G V N D R A W K Y G Y L D A D M N S N L R F
 6511 GCGCTGAGCA AGACGTTTCC CGTGAATAT GGCTCATAAC ACCCCTGTA TTACTGTAA TGTAAAGCAGA
 12 R S C S T E R Q I H S M
 6581 CAGTTTATAT GTCATGATG ATATATTTT ATCTTGTGCA ATGTAACATC AGAGATTTG AGACACAACG

43/43

6651 TGGCTTCCCC CCCCCCCCCA TTATTGAAGC ATTTATCAGG GTTATTGTCT CATGAGCGGA TACATATTTG
6721 AATGTATTTA GAAAATAAA CAAATAGGGG TTCCGCCAC ATTTCCCCCA AAATGCCAC CTGACGTCTA
6791 AGAAACCATT ATTATCATGA CATTAACTTA TAAAATAGG CGTATCACGA GGCCCTTTCG TCTCCGGCGT
6861 TTCTGGTGTAG ACGGTGAAAA CCTCTGACAC ATGCCAGCTCC CGGAGACGGT CACAGCTGTG CTGTAAAGCGG
6931 ATGCCGGGAG CAGACAAGCC CGTCAGGGCG CGTCAGCGGG TGTGGGGGG TGTGGGGCT GGCTTAACPA
7001 TGCGGCATCA GAGCAGATG TACTGAGAGT GCACCATATG CGGTGTGAAA TACCGCACAG ATCCGTAAGG
7071 AGAAAATACC GCATCAGATT GGCTATTGG